SCHOOL OF INFORMATION SCIENCES & TECHNOLOGY
DEPARTMENT OF STATISTICS
POSTGRADUATE PROGRAM

Some statistical models in ecology and epidemiology

By

Marios Kondakis

A THESIS

Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfilment of the requirements for
the degree of

Doctor of Philosophy in Statistics

Athens, Greece
September 2021

ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ
ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΔΙΔΑΚΤΟΡΙΚΟ ΠΡΟΓΡΑΜΜΑ

Στατιστικά μοντέλα στην οικολογία και την επιδημιολογία

Μάριος Κονδάκης

ΔΙΑΤΡΙΒΗ

Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Διπλώματος Διαδακτορικών Σπουδών στη Στατιστική

Αθήνα,
Σεπτέμβριος 2021

DEDICATION

To everyone who assisted in the completion of this work, both directly and indirectly, with selflessness and willingness.

ACKNOWLEDGEMENT

This dissertation, which was submitted to the Department of Statistics of the School of Informatics and Technology at Athens University of Economics and Business, would not have been possible without the uncompromising assistance of a number of people.

I'd like to take this opportunity to express to my supervisor committee. First and foremost, I want to express my heartfelt gratitude and appreciation to Prof. Nikolaos Demiris, who patiently and thoroughly oversaw the entire project, providing motivation, valuable advises and vast knowledge. He not only encouraged me in my research over the years, but he also allowed me to develop as a research scientist. Furthermore, I would like to express my sincere gratitude to Prof. Ioannis Ntzoufras, who always enriched our constructive discussions with his extensive knowledge and well-targeted and inspiring comments, which led to a broadening of the research vision to include a variety of topics. Furthermore, many thanks to Prof. Epaminonda Kyriakidi for his valuable suggestions and advice whenever it was required.

I am affiliated with Prof. Nikos E. Papanikolaou, who has been continuously assisting in the deeper understanding and interpretation of ecological processes, as well as advising, encouraging, and providing relevant experimental data. I would also like to thank all the members of my Dissertation Committee for their inspiring comments and suggestions.

In addition, I am grateful for the opportunity to work with a diverse group of people, especially in the Statistics department, each of whom contributed to the final product in their own unique way, and I feel compelled to express my gratitude to all of you—professors, PhD students, and friends.

No one was more important than my family in the creation of this work. There are no words to express how grateful I am to Theofilo, Maria, Loukilliano, Katerina, Gianni, Maki and Eugenia. Giota and young Theofilo Aggelo, for your love and all the sacrifices you made for me. This research project would not have been possible without the unwavering support of you.

ABSTRACT

This work focuses on statistically modelling specific biological processes from a Bayesian standpoint and it can divided into four components. The first component, is concerned with ecological models that describe the fitness of insects (mainly in the Phylum of arthropods), as described by deterministic and stochastic demographic models in order to understand population performance of invasive species, which can lead to management decisions, especially under the pressure of climatic change. The second component involves the investigation of non-linear statistical models based on popular ecological functions that describe the developmental process of arthropods as it is affected by temperature. Statistical modelling may provide insights into the population evolution of arthropod pests as well as biological control agents, which is important for ecology. Moreover, we investigate various computation techniques in order to not only derive robust estimates of the parameters of interest, but also to compare different models and computation methods. The third component entails modelling predator-prey systems to account for changes in prey population consumption over time as well as inter-individual interactions within the same species. Understanding how to describe and predict population performance, as well as how to manage invasive species, requires a thorough understanding of predator-prey population interactions. Hence we study statistical models that generate data using the Binomial distribution while prey density change in real time is described via ordinary differential equations (ode) ecological models. To address the possibility of noise, we propose that the probability of being consumed be linked to a stochastic process that is centered and reduced to (in the absence of diffusion) the instantaneous ratio of consumed prey density (which is the default link). The fourth section differs from the previous sections in that it focuses on modeling and detection of the spread of Vector-borne diseases (VBDs), as well as the development of a semi-automatic early warning system for the prevention of these diseases in the context of epidemiology. In particular, risk measures for the spread of malaria through mosquito bites are described, and proposals are presented for mapping R-tools illustrating the intensity of risk measures, based on actual data, allowing better decision making and prevention of dispersal through early warning. A generic observation running throughout this work is that detailed and robust modelling may assist greatly in more accurate and cautious conclusions drawn when interpreting the data.

ΠΕΡΙΛΗΨΗ

Η εργασία αυτή επικεντρώνεται στη στατιστική μοντελοποίηση συγκεκριμένων βιολογικών διεργασιών από τη Μπευζιανή σκοπιά και μπορεί να χωριστεί σε τέσσερα μέρη. Το πρώτο μέρος αφορά οικολογικά μοντέλα που περιγράφουν τη δυναμική των πληθυσμών των εντόμων (κυρίως των αρθροπόδων), όπως περιγράφεται από ντετερμινιστικά και στοχαστικά δημογραφικά μοντέλα, προκειμένου να κατανοήσουμε την πληθυσμιακή απόδοση των χωροκατακτητικών ειδών, η οποία μπορεί να οδηγήσει σε αποφάσεις διαχείρισης, ειδικά υπό την πίεση της κλιματικής αλλαγής. Το δεύτερο μέρος αφορά τη διερεύνηση των μη γραμμικών στατιστικών μοντέλων που βασίζονται σε δημοφιλείς οικολογικές συναρτήσεις που περιγράφουν την ανάπτυξη των αρθρόπόδων καθώς επηρεάζεται από τη θερμοκρασία. Η στατιστική μοντελοποίηση μπορεί να παρέχει πληροφορίες σχετικά με την εξέλιξη του πληθυσμού των αρθρόποδων, καθώς και των παραγόντων βιολογικής αντιμετώπισης, γεγονός σημαντικό για την οικολογία. Επιπλέον, διερευνούμε διάφορες τεχνικές υπολογισμού προκειμένου όχι μόνο να εξάγουμε αξιόπιστες εκτιμήσεις των παραμέτρων που μας ενδιαφέρουν, αλλά και να συγκρίνουμε διαφορετικά μοντέλα και μεθόδους υπολογισμού. Το τρίτο μέρος περιλαμβάνει την μοντελοποίηση συστημάτων θηρευτών-θηραμάτων ώστε να λαμβάνονται υπόψη οι αλλαγές στην κατανάλωση του πληθυσμού των θηραμάτων με την πάροδο του χρόνου καθώς και οι αλληλεπιδράσεις μεταξύ ατόμων εντός του ίδιου είδους. Η κατανόηση του τρόπου περιγραφής και πρόβλεψης των επιδόσεων του πληθυσμού, καθώς και του τρόπου διαχείρισης των χωροκατακτητικών ειδών, απαιτεί πλήρη κατανόηση των αλληλεπιδράσεων του πληθυσμού θηρευτών-θηραμάτων. Ως εκ τούτου, μελετάμε στατιστικά μοντέλα που παράγουν δεδομένα χρησιμοποιώντας τη διωνυμική κατανομή, ενώ η μεταβολή της πυκνότητας των θηραμάτων σε πραγματικό χρόνο περιγράφεται μέσω οικολογικών μοντέλων με συνήθεις διαφορικές εξισώσεις. Για την αντιμετώπιση πιθανού θορύβου, προτείνουμε η πιθανότητα κατανάλωσης να συνδεθεί με μια στοχαστική διαδικασία που είναι κεντραρισμένη και ίση (στην περίπτωση απουσίας της διάχυσης) με τη στιγμιαία αναλογία της πυκνότητας του θηράματος που καταναλώνεται (που είναι ο προεπιλεγμένος σύνδεσμος). Το τέταρτο μέρος είναι διακριτό από τα προηγούμενα και αφορά τη μοντελοποίηση και ανίχνευση της εξάπλωσης των ασθενειών που μεταδίδονται από φορέα, καθώς και με την ανάπτυξη ενός ημιαυτόματου συστήματος έγκυρης προειδοποίησης για την

πρόληψη των ασθενειών αυτών στο πλαίσιο της επιδημιολογίας. Ειδικότερα, περιγράφονται τα μέτρα κινδύνου για τη διάδοση της ελονοσίας μέσω τσιμπημάτων από τα κουνούπια, και παρουσιάζονται προτάσεις για τη χαρτογράφηση με εργαλεία στην R που απεικονίζουν την ένταση των μέτρων κινδύνου, με βάση πραγματικά δεδομένα, επιτρέποντας καλύτερη λήψη αποφάσεων και πρόληψη της διασποράς μέσω έγκαιρης προειδοποίησης. Μια γενική παρατήρηση που διέπει αυτή τη δουλειά είναι ότι η λεπτομερής και εύρωστη μοντελοποίηση μπορεί να βοηθήσει πολύ σε πιο ακριβή και προσεκτικά συμπεράσματα που συνάγονται κατά την ερμηνεία των δεδομένων.

Contents

List of Tables

List of Figures

LIST OF TERMS AND ABBREVIATIONS

ACT accelerated failure time 12, 13

ADVI automatic differentiation variational inference 2, 28, 80

AIC Akaike information criterion 13, 21, 22, 83

BIC Bayesian information criterion 83

BMA Bayesian model averaging 2, 28

CI confidence interval 11, 20, 21

Cr.I. credible interval(s) 21, 23, 49, 57, 58, 80, 82, 109

d.f. degrees of freedom 22

Dev deviance 22

DIC Deviance information criterion 83

EWS early warning system 90, 96, 97, 103, 104, 110, 111

HMC Hamiltonian Monte Carlo 2, 28, 29, 38, 39, 47–49, 51, 52, 56–60, 63, 65, 77, 80, 81, 83, 85–88, 107, 109

LooIC leave one out cross validation information criterion 83

MCMC Markov chain Monte Carlo 14, 26, 37, 39, 40, 69, 106

ode ordinary differential equations 2, 74, 77–80, 83, 85–88, 108

OU Ornstein-Uhlenbeck 74, 78, 83, 85, 86, 88, 108, 109

VBDs Vector-borne diseases 3, 90, 92, 100, 104, 110, 111

VBI Variational Bayes inference 2, 29, 38, 39, 51, 58

WAIC Watanabe information criterion 83

ZIP zero inflated Poisson 14, 24

Chapter 1

Introduction

I confirm that the work presented in this thesis is my own. Where information has been derived from other sources this has been clearly indicated throughout this thesis. Chapter 2 presents the statistical part of work that was published as (Papanikolaou et al. 2019). Chapter 3 presents the research that was submitted to the journal of Environmental and Ecological Statistics in August 2021 (Kondakis et al. 2021a). Chapter 4 describes the R-package created on GitHub at (Kondakis et al. 2021b) and will be submitted for publication in autumn 2021. Chapter 5 presents the statistical part of a paper that will be submitted for publication in autumn 2021. Chapter 6 provides the R-based spatial map tools that was the novelty in (Pergantas et al. 2021).

1.1 Demographic statistics

The study of demographic statistical methods and models, both deterministic and stochastic, provides useful insights into the fitness of economically important arthropods, such as the *Trogoderma granarium* (*T. granarium*) beetle, in order to describe the reproduction process, population increase, mortality, fecundity rates, survival time and other critical measures that assess population performance. Because demography is a primary domain of statistics, it provides a number of fundamental tools, including life tables, offspring modelling, and survival analysis parametric and non-parametric models. As a result, understanding population dynamics of invasive species, is important in ecology and can lead to management decisions, especially under the pressure of climatic change.

1.2 Developmental rates

The developmental rate of insects is another important feature in ecology. Insects, and thus arthropods, go through immature stages (called instars) in which their rigid exoskeleton cannot expand much and must be shed and replaced with a larger one (process called molting) as the insect grows. To that end, most insects' development includes their early-life transformation, from egg to adult. Insect species do not develop at the

same rate, and temperature is a major factor influencing their development. Hence, investigation of non-linear statistical models based on popular ecological functions that describe the developmental process of arthropods as it is affected by temperature fluctuations. Statistical modelling can provide insights into the population evolution of arthropod pests as well as biological control agents, which is critical for surveillance and containment of invasive species whose spread is accelerated by climate change, increased human population density, and international trade. When modelling developmental rates, several challenges arise. Some of them are that the non-linear mean structure in some ecological functions is truncated by parameters of interest, several model parameters have limited sampling space, and the presence of a specific data structure consisting of either positive reciprocals or zeros (indicating no development of the arthropod). Furthermore, comparing non-linear, non-nested models with varying parameter counts is not trivial. To address these problems, the Bayesian paradigm and, in particular, contemporary computation methods like HMC sampling that provides robust estimates even under these model and data structure conditions, as well as the Variational Bayes inference (VBI) methods for fast approximations to posterior distributions can be engaged. Similarly, Bayesian model averaging (BMA) techniques are useful so as to generate robust estimates for the parameters of interest by combining the predictive power of all models involved.

1.3 Predator-prey systems

The study of predator-prey population interactions is important in ecology for understanding how to describe and predict population performance, as well as how to manage invasive species. Modelling predator-prey systems is a critical tool in order to account for changes in prey population consumption over time while accounting for inter-individual interactions not only between species but also within the same species. To that end, it is natural to use the Binomial data generated scheme to model the number of prey consumed, as well as specific ode ecological models that describe the abundance of the prey density in real time. In addition, the presence of noise due to the nature of such ecological data, as well as the potential failure of models to detect any kind of systematic source of variation in such data, adds to the modelling challenges. As a result, involving stochasticity to the probability of prey being consumed by the predator is a way to deal with this kind of noise and derive better predictions. Information criteria are also useful to compare models not only between different ode methods, but also between deterministic and stochastic models. Posterior predictive plots alongside observed data points can depict how well the suggested models fit the data. The computation methods provided by the STAN software, can handle ode solution approximations within HMC sampling iterations, as well as within the automatic differentiation

variational inference (ADVI) methods.

1.4 Vector born diseases

Dengue fever, West Nile Virus, Lyme disease, and malaria are examples of VBDs caused by an infection spread to humans and other animals by blood-feeding arthropods like mosquitoes. As a result, statistical modelling of VBDs plays an important role in providing an explicit framework for understanding parasite disease transmission within and between hosts. The model class used in VBDs contains several directly calculated or no measures of the infected population that are derived each time by selecting clinical and biological information in a simplified form that appears to be important to the scientific question under investigation. Risk measures, among other things, provide insights into disease transmission dynamics and can be used as early detection tools for public awareness. The components for calculating risk measures, as well as the steps for using R-tools to create various types of spatial maps, can be combined to create a mapping tool that not only depicts potential infection points on a real-world map, but also the intensity of appropriate risk indicators. In terms of public health, the study's goal is to use the model class, as well as the associated geographical mapping of the statistical model's generated risk estimates on real data, as an early-warning system, allowing for better decision making and (VBD) prevention.

1.5 Overview of the Thesis

The effort in statistical modelling specific biological processes in this work can be divided into four components, the first three of which focus on modelling of ecological processes that describe the reproducing rate, mortality rate, developmental rate and consuming rate of insects (in the Phylum of arthropods), while the fourth component focuses on epidemiological methods for calculating risk measures of disease transmission from vectors (such as mosquitoes) and depicting them spatially on maps for warning purposes. The structure of this work, matching the aforementioned components with the following Chapters is: The first component which is concerned with reproducing and mortality dynamics is addressed in Chapter 2, the second with developmental dynamics is addressed in Chapters 3 and 4, the third with predator-prey dynamics in Chapter 5, and the fourth with VBDs transmission dynamics in Chapter 6.

Chapter 2

Demographic analysis of ecological data

2.1 Introduction

In this chapter, we use deterministic and stochastic demographic tools to elucidate the fitness components of invasive species such as *T. granarium*, providing critical information for their management. This work was published in the journal PLOS One in 2019 as (Papanikolaou et al. 2019). The viability of the populations of living organisms is strongly dependent on their fitness, referring to their ability to survive and reproduce in a specific environment (Orr 2009). Survival and reproduction are critical aspects of population dynamics, regulating their growth rate and allowing for several temporal fluctuations (Huey and Berrigan 2001*a*, Engen et al. 2009, Compagnoni et al. 2016). To this end, ecologists are often interested in understanding the patterns of these biological features in order to describe and predict populations' performance (Hunter et al. 2010, Jonzén et al. 2010, Stark and Banks 2016). Insects such as ectotherm organisms are characterised by the fact that their body temperature converges to the one of the environment they are exposed to (Norris and Kunz 2012). This affects the rate of metabolism, the biochemical reactions which facilitate production and energy release, as well as the synthesis of necessary molecules that serve as structural or functional components (Neven 2000, van der Have 2008). In particular, temperature affects the functionality of enzymes, which in turn act as catalysts for these biochemical reactions (van der Have 2008). Consequently, within a range of temperatures in which insects develop and reproduce, various biological features are affected, such as mortality, reproduction, life span, and growth rate (Huey and Berrigan 2001*a*, Berven 1990, Nedvěd 2009, Papanikolaou et al. 2013, 2014). Thus, the performance of the insects is subject to several temporal fluctuations in terms of population size through time. Understanding populations' performance is of particular importance, as their assessment can lead to decisions on their management (Hare et al. 2011), particularly under the pressure of climatic change (Bradshaw et al. 2016). Hence, elucidating into fitness components can lead to a clearer understanding of an organism's contribution to future generations and therefore its potential population development (Huey and Berrigan 2001*a*, Jonzén et al. 2010, Polanco et al. 2011). Demography represents the standard tool used for elu-

cidating the fitness components of living organisms, as it allows for an integrated and comparative description of several biological processes, as well as an investigation on the organisms' mortality and reproduction patterns (Carey 2001). Assuming a closed population (i.e., no migration) with stable age distribution, applied demography allows for the calculation of several population parameters, tabulating the birth and death rates of the organisms of interest in a cohort life table (Carey 1993). Among these, the two most commonly used measures of fitness are the intrinsic rate of increase (r_m), which represents the implicit rate of population increase independent of initial age structure, and the net reproductive rate (R_o), which is typically interpreted as the average number of female offspring that a female gives birth to over her lifetime (Huey and Berrigan 2001a). The invasive dermestid khapra beetle, *Trogoderma granarium*, is an economically important stored-product species that is subject to strict phytosanitary measures (Banks 1977, Lowe et al. 2000, Hill 2002, EPPO 2013). Native to India (Rahman et al. 1945), its host range now includes Africa, Asia and Europe (Aitken 1975, Peacock et al. 1993, Athanassiou et al. 2015). *T. granarium* is categorized as an A2 quarantine organism (EPPO 1981, 2013, 2018), as it is under quarantine regulation in numerous countries (EPPO 2018). The rapid increase in interceptions at US ports is a cause of concern (Hagstrum et al. 2012). This trend is also evident in Europe considering the interceptions that have been recorded in numerous countries, mostly in central Europe, including Austria, Bulgaria, Croatia, Czech Republic, Italy, Poland, Portugal and Slovakia (EPPO 2018). Despite the economic importance of *T. granarium*, there are no data on the demography of this species at different temperatures, which could provide valuable knowledge on its outbreaks and expansion and thus timely and effective management. In deterministic demographic models, the output of the model is fully determined by the parameter values and the initial conditions. On the other hand stochastic demographic models naturally quantify the randomness that stems from the inherent variability of the population and also allow for model assessment and exploration of the appropriate probability distribution for each element of interest. The latter serves towards our broader aim of embedding the current study within the stochastic approach to demography. This viewpoint has turned our attention to the stochastic modelling of the survival time (time until an event of interest occurs such as the lifespan) of *T. granarium* and testing for statistically significant differences with respect to temperature treatments. We also extensively investigated the distribution of the time to the first birth, an event of primary interest. Our data contained a number of *T. granarium* beetles which gave no birth during their lifetime. Therefore, in contrast to the observations concerned with time-to-death, the time to the first birth data contained censored observations, necessitating a survival type of analysis. In addition, measuring the number of offspring of *T. granarium* beetles on a daily basis results in the observation of an excessive number of zeros. The zero-inflated-Poisson model was fitted using a Bayesian

approach, leading to accurate point and interval estimates, even in the presence of a large percentage of zeros in the sample (Ghosh et al. 2006). The output of deterministic and stochastic demography is combined in chapter 2 to investigate the survival or extinction of *T. granarium*, as well as the patterns of mortality and reproduction, and to provide a comparative calculation of its demographic parameters in different temperature ranges. In order to calculate confidence limits and perform hypothesis testing across temperature ranges, we specifically introduce uncertainty into deterministic demographic cohort measures described by formulae (2.8-2.20). Additionally stochastic models described in formulae (2.24), (2.25), (2.26), (2.27) and (2.28) are used in conjunction with survival analysis techniques defined in formula (2.22) to investigate both the time until death and the time until the first egg is produced. Furthermore, the number of offspring is modeled using the Bayesian paradigm and probability density in (2.30), with excess zeros taken into account. Finally, we search at how the fitness components of *T. granarium* used in all of the preceding approaches change with temperature. To that end, stochastic models, while underutilized in demographic studies, could provide important information on *T. granarium* functionality.

2.2 Deterministic demographic approaches

Demography is derived from the Greek root demos (people) and literally means 'description of the people'. In 1855, Achille Guillard defined demography as 'the natural and social history of the human species or the mathematical knowledge of populations, their general changes, and their physical, civil, intellectual, and moral condition'. The methodology of demographic studies includes data collection, demographic analysis, and data interpretation. Individuals, populations, cohorts, and demographic rates are some of the key concepts and tools addressed by demographic analysis. Individuals are defined as "single organisms that are carriers of demographic attributes" by Frans Willekens in 1986, and some basic characteristics of individuals are the developmental rate, the age-specific level of reproduction, and the time until death. The population is broadly as "a group of individuals coexisting at a given moment" (Pressat 1985). The cohort is defined as "a group who experience the same significant event in during a specific time period, and who can thus be identified as a group for subsequent analysis" (Pressat 1985). Cohort attributes are commonly explored using demographic parameters that form a life table in the form of age schedule events. Cohorts determine population traits in this way. A complete cohort life table includes the mortality experience of a specific cohort from the moment of birth until no individuals remain in the original cohort. The demographic rates are classified into five groups based on the type of population counted in the denominator or the type of events counted in the numerator. The crude rates, in particular, are applied to all individuals rather than dividing them

by age or sex groupings, the age-specific rates, which are the same as crude rates but with age restrictions, the restricted rates, which are applied to any special sub-group, the rates by topic, which apply to each specialized topic in demography such as the total fertility rate, the gross reproductive rate, or the net reproductive rate and the intrinsic rates which prevail in a stable population and are invariant to any accidental or transient short-term feature of the age distribution.

There are two fundamental approaches to describe the population rate of change that take birth and death rates into account. First, the crude rate model is introduced, which assumes that i) the population is not structured by age (homogeneity assumption), ii) birth and death rates remain constant, and iii) the population is closed (i.e., no migration). Second, assuming the last two assumptions are true, the stable population model (Lotka's equation) is developed, along with a common age structure, which adds a more realistic and interesting dimension while not changing the geometric manner of growth introduced by the crude rate model.

Despite the fact that it considers change at a constant rate forever, the crude rate model is fundamental to demography because it defines population change as a compounding process, it establishes a foundation for examining population growth patterns over short time periods, and it provides the initial framework from which to build more complicated models (Carey 1993). It is based on the following balancing equation which connects the total population at time t (counted in years, months, weeks, days and so on) to the total population at time $t-1$:

$$N_t = N_{t-1} + births_t - deaths_t + immigrants_t - emigrants_t, \tag{2.1}$$

where $births_t$, $deaths_t$, $immigrants_t$, $emigrants_t$ are the births, the deaths, the number of immigrants and emigrants respectively at period from $t-1$ until t. For simplicity migration in (2.1) is typically neglected in demographic analysis and the number of births and deaths are substituted by the products:

$$\begin{aligned} births_t &= b \cdot N_{t-1}, \\ deaths_t &= d \cdot N_{t-1}, \end{aligned}$$

where b and d are the per ca-pita birth and death rates respectively. Denoting N_0 the initial population at the beginning of the study at time 0, the population at time t is given by

$$\begin{aligned} N_t &= N_0 \cdot (1 + b - d)^t, \\ N_t &= N_0 \cdot (\lambda)^t. \end{aligned} \tag{2.2}$$

When $\lambda > 1$, the population grows, whereas when $\lambda < 1$, the population shrinks.

The population can now be calculated using the natural logarithm in equation 2.2 and a new parameter $r = ln(\lambda)$:

$$N_t = N_0 \cdot e^{r \cdot t}, \tag{2.3}$$

where r is the intrinsic rate of increase, which was introduced by Dublin and Lotka in 1925 as a measure of the rate of natural increase and is interpreted as the constant exponential growth rate per individual per time unit.

Alfred Lotka developed the stable population model in 1907 by expanding the crude rate model and adding a structure to the population's age. The latter, in particular, identifies the two most important population parameters (the age distribution and the growth rate), their interdependence, and their relationship with the birth and death cohort parameters. It is capable of producing actual numbers that explain the patterns of the rate of growth, the number of people in each age group, and the population size at a given time (Carey 1993). In the simplest two-class age structure of the population, the number of people at age x_0 and age $x_0 + 1$ at time $t + 1$ is given by the following system of equations:

$$\begin{aligned} N_{x_0,t+1} &= m_{x_0} \cdot N_{x_0,t} + m_{x_0+1} \cdot N_{x_0+1,t}, \\ N_{x_0+1,t+1} &= p_{x_0} \cdot N_{x_0,t}, \end{aligned} \tag{2.4}$$

where m_{x_0} and p_{x_0} are the age-specific birth rate and the probability of survival from birth age x_0 to age $x_0 + 1$, respectively. Commonly, birth age x_0 is taken as zero. The rate of increase λ is constant between age classes given by:

$$\lambda = \frac{N_{x_0+1,t+1}}{N_{x_0+1,t}} = \frac{N_{x_0,t+1}}{N_{x_0,t}} \tag{2.5}$$

Combining the formulae in 2.4 and 2.5 and dividing by λ^2 yields equation (2.6), which can be generalized in Lotka's equation (2.7).

$$0 = N_{x_0,t} \cdot \left(-1 + m_{x_0} \cdot \lambda^{-1} + m_{x_0+1} \cdot p_{x_0} \cdot \lambda^{-2}\right) \tag{2.6}$$

$$1 = \sum_{k=0}^{\omega} \left\{\lambda^{-(k+1)} \cdot l_k \cdot m_{x_0+k}\right\}, \tag{2.7}$$

where age k ranges from 0 to a positive number ω and l_k represents cohort survival to age $x_0 + k$. The parameters of interest used in the current study are presented in the following paragraph 2.2.1.

2.2.1 Demographic parameters of interest

Following (Carey 1993, Kontodimas et al. 2008, Papanikolaou et al. 2014, Zeki et al. 2015), the demographic parameters of interest used to describe the insects' population dynamics are the following.

- The cohort survival to age x:

$$(l_x), \tag{2.8}$$

  which is equal to the proportion of individuals that have survived from birth to age x. The number of individuals survived is usually associated to an arbitrary constant known as the life table radix (Carey 1993). In population biology the radix is set to unity so that the rate measures used are expressed as a fraction to the original number of individuals. The initial time is usually taken as zero and $l_0 = 1$.

- The age specific mortality:

$$q_x = 1 - \frac{l_{x+1}}{l_x}, \tag{2.9}$$

  which represents the probability of dying over period (x,x+1].

- The age specific fecundity:

$$(m_x), \tag{2.10}$$

  which represents the averaged number of offspring produced by an individual at age x (reproductive age schedule) and is calculated by multiplying the mean number of eggs per female at age x by the female-to-total-population ratio (observed by sorting 100 offspring).

- The net reproductive rate:

$$R_0 = \sum_{x=0}^{\omega} (l_x \cdot m_x), \tag{2.11}$$

  which represents the per capita rate of offspring production over a time equal to cohort study period from $x = 0$ to $x = \omega$.

- the finite rate of increase:

$$(\lambda), \tag{2.12}$$

  which represents the per capita rate of population growth in each time step according to (2.5).

- The intrinsic rate of increase:

$$(r_m = ln\,(\lambda)), \tag{2.13}$$

which represents the per capita exponential rate of increase in a closed population (that has been subjected to constant age-specific schedules of fertility and mortality for a long period) in each time step. Using formulae (2.13) and (2.7), the intrinsic rate is calculated as the solution of the equation:

$$1 = \sum_{x=0}^{\omega} \left\{ e^{-r_m \cdot (x+1)} \cdot l_x \cdot m_x \right\}, \tag{2.14}$$

which is the discrete time Lotka's equation. If the age variable is continuous, the equation (2.14) is as follows:

$$1 = \sum_{x=0}^{\omega} \left\{ e^{-r_m \cdot x} \cdot l_x \cdot m_x \right\}. \tag{2.15}$$

- The mean generation time:

$$\tilde{T} = \frac{\sum_{x=0}^{\omega} (x \cdot l_x \cdot m_x)}{\sum_{x=0}^{\omega} (l_x \cdot m_x)}, \tag{2.16}$$

$$T = \frac{lnR_0}{r_m}, \tag{2.17}$$

where $\tilde{T}$ in (2.16) represents the cohort mean age of reproduction, which characterizes the mean interval between births of one generation and those of the next (Pressat 1985). T in (2.17) also represents the mean generation time, which is defined as the time required for the population to increase by a factor equal to the net reproductive rate. Formula (2.17) is derived by substituting the mean generation time T in formula (2.13). In the discrete age case, the difference between formulae (2.14) and (2.17) gives that $\tilde{T} = T + 1$.

- The doubling time:

$$DT = \frac{ln2}{r_m}, \tag{2.18}$$

which represents the time required for the population to double. It is derived if $\lambda^t = 2$ in the formula (2.3).

- The reproductive value of the females:

$$V_x = \frac{\sum_{y \geq x} (e^{-r_m \cdot y} \cdot l_y \cdot m_y)}{e^{-r_m \cdot x} \cdot l_x}, \tag{2.19}$$

which represents the contribution of an individual of a given age x to the future population growth.

- The expected remaining life time at age x:

$$E_x = \frac{\sum_{t \geq x} \frac{l_t + l_{t+1}}{2}}{l_x}, \tag{2.20}$$

which represents the expected remaining life time of an individual at age x.

The aforementioned parameters were estimated at 30, 35 and $40°C$, 65% relative humidity and continuous darkness for the real-life data example in Section 2.4. Significant differences between life table parameters at each of the examined temperature were tested via a Wald test, essentially the superposition of 95% confidence intervals (CIs). This is a general method for hypothesis testing and avoids the recent controversy with the use of p-values in the statistical literature (Halsey 2019). The CIs were obtained by bootstrapping (Efron 1992) in R (Team 2021), sampling with replacement 1000 datasets in each temperature group and re-estimating the parameters for each set. This technique avoids unnecessary asymptotic normality assumptions and estimates the CIs using the empirical 2.5% and 97.5% percentiles, yielding general and robust procedures for statistical estimation and hypothesis testing.

2.2.2 Modelling temperature-dependent intrinsic rate of increase

The relationship between temperature and the intrinsic rate of increase was described by the Briere model (Briere et al. 1999), which is of the form:

$$r(T) = \alpha \cdot T \cdot (T - T_{min}) \cdot (T_{max} - T)^{\frac{1}{2}}, \tag{2.21}$$

where T denotes the ambient temperature, α is an estimated parameter, T_{min} is the lower and T_{max} the higher temperature in which the intrinsic rate of increase is equal to zero. We proceeded by assuming that at the temperature of 17.20°C the intrinsic rate of increase is equal to zero since no development has been detected (Burges 2008) in this temperature. The limited capacity of deterministic models in predictions concerned with alternative environmental conditions or the sensitivity of the beetles' biochemical reactions to other environmental conditions (Burg 2014) like pressure and substances of air breathing, suggest that the randomness of the mechanism based on which the *T. granarium* female beetles incubate can be appropriately modeled by a stochastic process as opposed to a deterministic model and this is described in the following subsection.

2.3 Stochastic demographic approaches

2.3.1 Kaplan-Meier method

Survival analysis techniques were utilized in order to examine (in a life cycle generation) both (i) the time (in days) until the event of "death" and (ii) the time (in days) until the event of *T. granarium* become active. In both events of interest, the time until the event occurs can be considered as a non-negatively-distributed random variable (Cox and Oakes 1984). In order to compare the survival times of *T. granarium*, kept at different incubators the predictor variable was the temperature level at 30, 35 and $40^{\circ}C$ respectively. The survival probabilities of *T. granarium* for each temperature group were estimated using the Kaplan-Meier product-limit estimator (Kaplan and Meier 1958). The Kaplan–Meier is a nonparametric maximum likelihood estimator of the survival function "$S(t)$" and is given by the product:

$$\hat{S}\left(t\right)=\prod_{i:t_i\leq t}\left(1-\frac{d_i}{n_i}\right), \tag{2.22}$$

where the observed survival times t_i are sorted in ascending order $t_1 < t_2 < \ldots$, the parameter d_i represent the number of events at time t_i while, parameter n_i is the number of the individuals known to survive at time t_i. The Kaplan–Meier estimator is one of the most frequently used methods of survival analysis and can be used to examine the effectiveness of treatment or the influence of a factor in the survival times of the insects under study (Kaplan and Meier 1958).

2.3.2 Parametric Survival function models

In addition, some commonly used distributions in survival analysis were considered to fit a parametric survival regression model to the *T. granarium* data. Specifically, the distributions used were: (i) the Exponential (2.25), (ii) the Weibull (2.26), (iii) the Lognormal (2.27) and (iv) the Log-logistic (2.28). The parametric survival regression used to fit the *T. granarium* data is the accelerated failure time (ACT) model (Wei 1992, Kalbfleisch and Prentice 2011) which is a parametric model of the form

$$S(t|\theta)=S_0(e^{X'\theta}t), \tag{2.23}$$

where S_0 is a function for the baseline survival rate (which is related to the mean survival time). The ACT model uses covariates in term $e^{x'\theta}$ to place individuals on different time scales. The term $e^{X'\theta}$ is called the acceleration factor as it scales (accelerates) the survival time for each covariate . The ACT model can be rewritten in a log-linear form:

$$logT = \mu + X'\theta + \sigma \cdot W, \tag{2.24}$$

where the logarithm of survival time $(logT)$ is linearly related to its mean μ, to the acceleration factor $X'\theta$, and to an error term W scaled by a parameter σ.The variable W describes the error distribution and is chosen so that T is one of four well-known distributions, the probability density functions of which are given below:
the Exponential with rate parameter λ

$$p(t) = \lambda\, e^{-\lambda t}, \tag{2.25}$$

the Weibull with shape parameter a and scale parameter b

$$p(t) = \frac{a}{b}\left(\frac{t}{b}\right)^{a-1} e^{-(t/b)^a}, \tag{2.26}$$

the Lognormal with mean μ and std. deviation σ of the log(T)

$$p(t) = \frac{1}{\sigma t\sqrt{2\pi}} e^{\frac{-(logt-\mu)^2}{2\sigma^2}}, \tag{2.27}$$

the Log-logistic with shape parameter a and scale parameter b

$$p(t) = \frac{\frac{a}{b}\left(\frac{t}{b}\right)^{a-1}}{\left(1+\left(\frac{t}{b}\right)^a\right)^2} \tag{2.28}$$

The explanatory variable used in all the parametric models was the temperature. In order to fit the ACT model to the *T. granarium* data, the "survreg()" function from the survival package in R (Team 2021) is used.

Selection of the most suitable model was based on the Akaike Information Criterion (Sakamoto et al. 1986). The AIC given in (2.29) is a composite measure accounting for the goodness of fit of each model to the observations via the deviance, penalised for the model's complexity by adding twice the number of estimated parameters k. The AIC is an estimator of the relative quality of statistical models for a given set of data. It is defined as:

$$AIC = -2\, log(\hat{L}) + k, \tag{2.29}$$

Where $log(\hat{L})$ is the logarithm of the likelihood of the data given the current model and k is the number of estimated parameters. Its value is derived for each model by either calculating the $log(\hat{L})$ and k manually using the results of the "survreg()" function or by using directly the "extractAIC()" or the "AIC()" function in the survival package in R (Team 2021).

2.3.3 Modelling egg-counts with an excessive number of zeros

The statistical analysis of the number of eggs of *T. granarium* beetles was based upon the zero-inflated class of models. Herein, the underlying distribution considered for egg counts was the Poisson distribution leading to the zero inflated Poisson (ZIP) model (Ghosh et al. 2006). We used Bayesian methods for estimating the model parameters through the WinBUGS package (Lunn et al. 2000), a general purpose software designed to run Markov chain Monte Carlo (MCMC) simulations for a wide range of Bayesian models. The output of the ZIP model was obtained by running the WinBUGS software for 7000 samples within each temperature group, having 4000 iterations as burn-in. Specifically, burn-in refers to the initial, potentially non-stationary, portion of the Markov Chain measured in number of iterations. It pertains to the practice of discarding a number of samples at the start of the MCMC algorithm in order to allow the Markov Chain to reach its equilibrium (stationary) density which corresponds to the posterior distribution of interest.

The ZIP model in (2.30) used is a mixture model for each of the datasets corresponding to the three different temperature levels of 30, 35 and $40°C$ respectively. Let us denote with y_{ij} the response variable, the number of eggs laid by the i^{th} *T. granarium* beetle in the j^{th} day, so that j corresponds to the lifetime of the i^{th} beetle from its entry in the study until its death. Then, the statistical model we used posits that y_{ij} is either zero, with probability p, in which case no eggs are laid from i^{th} beetle in j^{th} day, or follows a Poisson distribution with parameter λ whence the i^{th} beetle generates an average of λ eggs in j^{th} day.

$$P(y_{ij}|p,\lambda) = \begin{cases} p + (1-p)\cdot e^{-\lambda} \text{ if } y_{ij} = 0 \\ (1-p)\cdot \frac{\lambda^{y_{ij}}\cdot e^{-\lambda}}{y_{ij}\,!} \text{ if } y_i \neq 0 \end{cases} \tag{2.30}$$

The model is completed with vague priors, namely a Uniform density on the (0, 1) interval for p and a Gaussian with mean zero and large variance on $log(\lambda)$.

2.4 *Trogoderma granarium* beetle example

2.4.1 The invasive *Trogoderma granarium*

The invasive dermestid khapra beetle, *T. granarium*, is an economically important stored-product species that is subject to strict phytosanitary measures (Banks 1977, Lowe et al. 2000, Hill 2002, Hagstrum et al. 2012, EPPO 2013). Native to India (Rahman et al. 1945), its host range now includes Africa, Asia and Europe (Aitken 1975, Peacock et al. 1993, Athanassiou et al. 2015). *T. granarium* is categorized as an A2 quarantine organism (EPPO 2013, 2018, 1981), as it is under quarantine regulation in numerous countries (EPPO 2018). The rapid increase in interceptions at US ports is a cause of

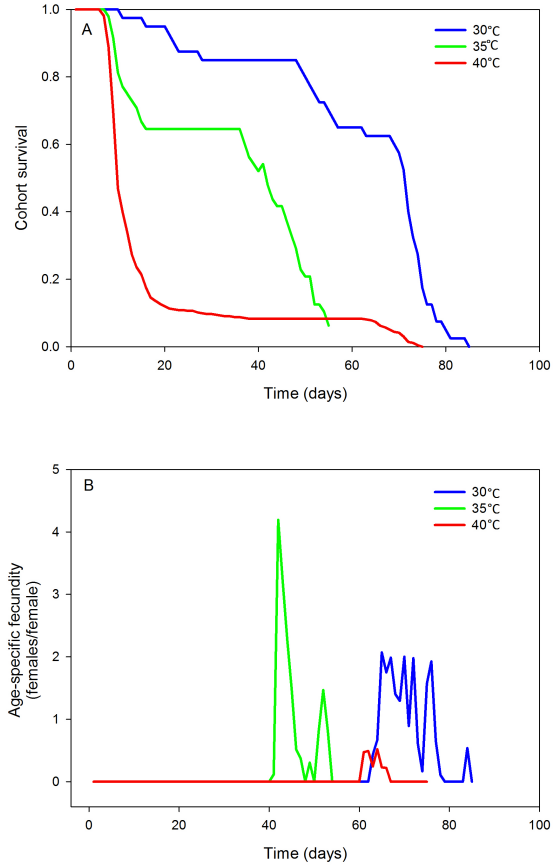

Figure 2.1 Plot of the cohort survival (A) and the age-specific fecundity (B) of *T. granarium* at constant temperatures.

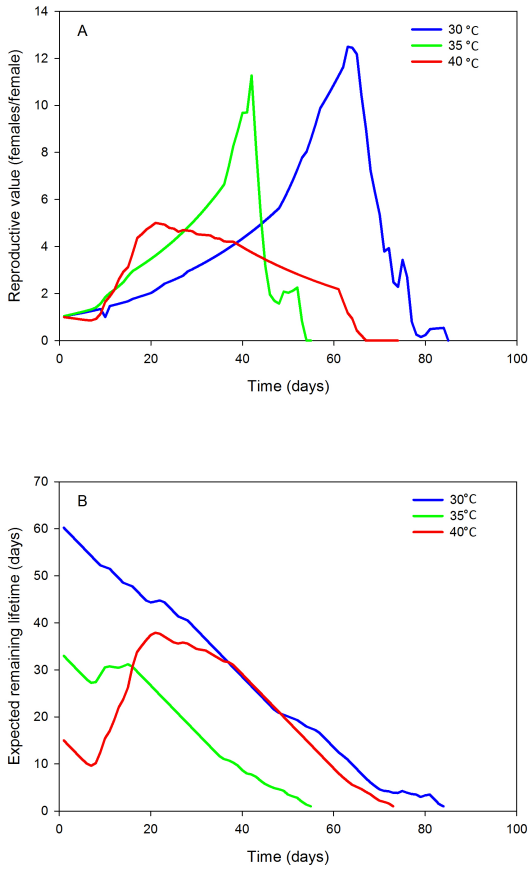

Figure 2.2 Plot of the reproductive value (A) and the expected remaining life time (B) of *T. granarium* females at constant temperatures.

concern (Hagstrum et al. 2012). This trend is also evident in Europe considering the interceptions that have been recorded in numerous countries, mostly in central Europe, including Austria, Bulgaria, Croatia, Czech Republic, Italy, Poland, Portugal and Slovakia (EPPO 2018).

2.4.2 Experimental design

The real-data used are gathered from a clinical trial made in lab. *T. granarium* were reared on wheat at $30°C$, 65% relative humidity at continuous darkness. The insect colony was established in 2014 from insects collected in Greek storage facilities in central Thessaly and southern Attica and since then it has been kept at the Laboratory of Agricultural Zoology and Entomology of the Agricultural University of Athens. In all experiments we used pesticide-free wheat (*Triticum durum*, var. Claudio) in order to maintain insect colonies. The moisture content of wheat was 12.1%, as determined by a calibrated moisture meter (mini GAC plus, Dickey-John Europe S.A.S., Colombes, France) at the beginning of the tests. Samples of 1 g of cracked wheat were separately put inside each Petri dish (8 cm diameter, 1.5 cm height). Wheat was cracked in a hand-

mill. Two testing sieves were used to make particles of cracked wheat approximately consistent. First, the cracked wheat was sieved with a No 30 (2.36 mm openings) US standard testing sieve (Advantech Manufacturing, Inc., New Berlin, WI). Subsequently, the sifted material was sieved again with a No 10 (2.00 mm openings) US standard testing sieve (Retsch GmbH, Haan, Germany). Then, the content of the latter was used for experimentation. The quantities of 1 g were weighed with a Precisa XB3200D compact balance (Alpha Analytical Instruments, Gerakas, Greece). The closures of the dishes bore a 1.50 cm diameter circular opening in the middle that was covered by muslin gauze to allow the sufficient aeration inside the dish. The upper inner walls of the dishes were covered by polytetrafluoroethylen (60 wt $\%$ dispersion in water) (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany) to prevent the escape of larvae and adults. To obtain eggs of *T. granarium*, 50 unsexed adult individuals, approximately 7 d old, were transferred from the culture to a 250 ml glass jar that contained 125 ml white soft wheat flour for 1 day. Then, the adults and eggs were separated from the flour with a No 20 and a No 60 U.S. standard testing sieves (Advantech Manufacturing, Inc., New Berlin, WI). The eggs that were remained on the mesh openings of the sieve were put in a Petri dish and inspected daily at 57 times total magnification of an Olympus stereomicroscope (SZX9, Bacacos S.A., Athens, Greece). Totally, 40, 48 and 433 eggs were used to obtain egg to adult development and mortality at 30, 35 and $40^{\circ}C$, respectively. We used higher number of eggs at $40^{\circ}C$ due to detrimental impact of this temperature to *T. granarium* survival. Newly hatched *T. granarium* larvae were very carefully separately placed inside each dish, with a fine brush (Cotman 111 No 000, Winsor and Newton, London, UK), that contained the cracked wheat. The dishes were placed in incubators set at the respective temperature and 65% relative humidity during the entire experimental period. The duration and survival of egg, larval, and pupal stages were recorded every 24 hours. In addition, female longevity and fecundity were examined daily. Formed pairs were kept separately in Petri dishes. We used 25, 26 and 36 pairs at 30, 35 and $40^{\circ}C$, respectively. The insects' thermal window, i.e. the range in temperature between the minimum and maximum rate of development for individual species, is about 20 °C (Dixon et al. 2009). Considering also that below $30^{\circ}C$ larvae of *T. granarium* fall to diapause prolonging their life up to 8 years (Aitken 1975, Peacock et al. 1993), and that *T. granariun* prefers environments with elevated temperatures (Banks 1977, Howe 1965), we selected 30, 35 and $40^{\circ}C$ as the most suitable temperature range for our study. It should be noted that during the summer, air temperature in Greece may reach or potentially exceed $40^{\circ}C$.

Table 2.1 Values of net reproductive rate (R_0), intrinsic rate of increase (r_m), finite rate of increase (λ), mean generation time (T) and doubling time (DT) of *T. granarium* reared on wheat (mean, 95% Confidence Intervals) at constant temperatures.

Temperature	$30°C$	$35°C$	$40°C$
Net reproductive rate R_0 (females/female)	9.73 (4.09, 16.28)	6.55 (2.36, 11.31)	0.17 (0.07, 0.30)
Intrinsic rate of increase r_m (females/female/d)	0.03 (0.02, 0.04)	0.04 (0.02, 0.06)	-0.03 (-0.05, -0.02)
Finite rate of increase λ	1.03 (1.02, 1.04)	1.04 (1.02 ,1.06)	0.97 (0.96, 0.98)
Mean generation time (d) T	69.04 (66.02, 71.65)	42.96 (41.22, 44.65)	58.12 (55.86, 61.48)
Doubling time (d) DT	22.2 (17.09, 34.37)	17.73 (12.26, 33.89)	-23.08 (-33.05, -15.42)

2.4.3 Demographic parameters results

The estimated demographic parameters showed considerable variation across the different temperature regimes used in this study. This was evident based upon the inspection of the 95% confidence intervals (Table 2.1) which were also used for hypothesis testing. Thus, we test for statistical significance using a 5% significance level.

The net reproductive rate did not differ significantly at 30 and $35°C$, but was substantially lower at $40°C$. The same trend was also established for the values of the intrinsic and the finite rate of increase as temperature increased from 30 and $35°C$. In contrast, the corresponding values at $40°C$ where close to zero. The doubling time also did not differ significantly at 30 and $35°C$ while a significantly lower doubling time was estimated at $40°C$. The mean generation time differed in all three temperatures, being significantly longer at $30°C$, shorter at $35°C$ and an intermediate estimate at $40°C$. The cohort survival decreased through time as presented in Fig. 2.1, while the age-specific fecundity increased until a particular age-dependent temperature, where a subsequent decrease follows (Fig. 2.1). In addition, females of approximately 63, 42 and 21 day-old reach their maximum reproductive potential at 30, 35 and $40°C$, respectively (Fig. 2.2). The expected remaining life time of *T. granarium* females at 30, 35 and $40°C$ is depicted in Fig. 2.2, reflecting that the initial decrease in this parameter is followed by an increase (although marginally at $30°C$) and an ultimately decrease. The p-values for testing the hypotheses that the demographic parameters are zero were smaller than 0.01.

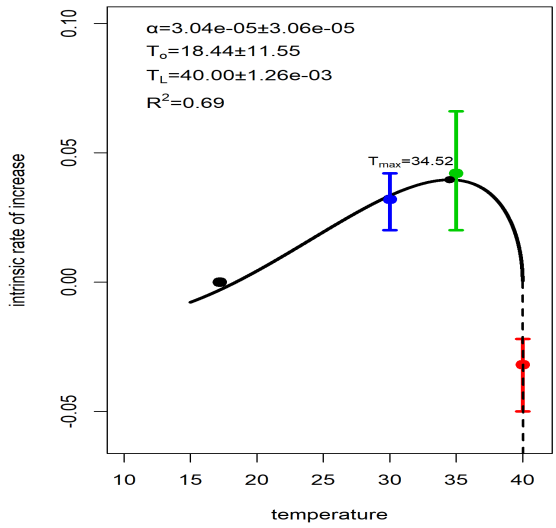

Figure 2.3 Estimated parameters and fitting of the Briere model to the intrinsic rate of increase data of *T. granarium*. Pointwise 95% C.I. are also depicted for each mean at $30°C$, $35°C$ and $40°C$ respectively.

2.4.4 Intrinsic rate results

The Briere model fitted reasonably well (R^2 is equal to 0.69 and standard error of the regression equal to 0.03, Fig. 2.3) to the intrinsic rate of increase data of *T. granarium*. The estimated minimum and maximum temperatures where the intrinsic rate of increase is expected to reach zero are 18.44 and 40.00°C respectively, obtaining its maximum value at 34.52°C.

Table 2.2 Estimated parameters of the Briere model fitted to *T. granarium* intrinsic rate of increase data.

Parameter	
α	$3.04e-5 \pm 3.06e-5$
T_0	18.44 ± 11.55
T_L	$40.00 \pm 1.26e-3$
R^2	0.69

2.4.5 Stochastic models results

2.4.5.1 Non parametric models

We modeled the times (in days) to two distinct types of event, the time until death, in which case there is no censoring due to the experimental design and the time until laying the first egg. The latter is subject to right censoring since some beetles die before they ever lay any egg. Hence, the median and other functionals of the survival times

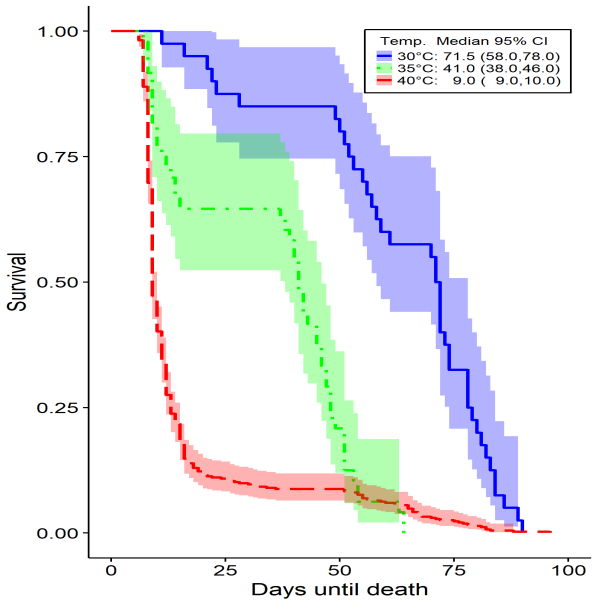

Figure 2.4 Kaplan-Meier survival curve estimator (Y-axis) vs. survival times (X-axis) of *T. granarium* until death along with their 95% confidence bands at $30°C$, $35°C$ and $40°C$ respectively. No censored observations exist due to the experimental design. In the legend we report the median and its 95% C.I.

are affected due to censoring which is substantial for the beetles studied at $40°C$. The survival times and their 95% confidence intervals are derived using the Kaplan-Meier estimators for the different temperature levels at 30, 35 and $40°C$. The results are depicted on Figs 2.4 and 2.5 respectively. The means of the survival time until death of *T. granarium* are 62.9, 34.2 and 15.6 days while their medians diminish rapidly (Fig. 2.4 and Table 2.3). Furthermore, the means of the time until first egg release are 71.7, 46.9 and 81.3 days while the medians decrease when the temperature rises from 30 to $35°C$ (Fig. 2.5 and Table 2.4). At $40°C$ it is apparent (Fig. 2.5) that the probability of *T. granarium* laying the first egg does not cross the 0.5 line and therefore the median time to first birth cannot be estimated.

Table 2.3 Survival times (mean, std. error, median and 95% CI) until death of *T. granarium* at constant temperatures.

Survival estimates until death		
Temperature	Mean $\pm$ se	Median and 95%CI
$30°C$	62.9 ± 3.4	71.5 (58, 78)
$35°C$	34.2 ± 2.7	41 (38, 46)
$40°C$	15.6 ± 0.8	9 (9, 10)

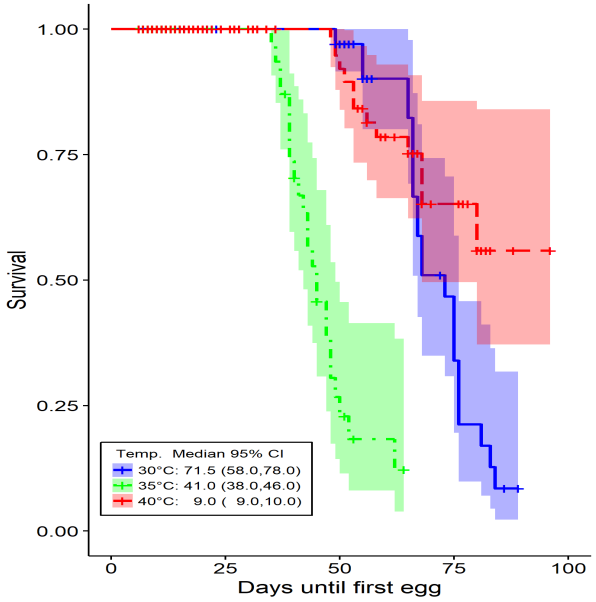

Figure 2.5 Kaplan-Meier survival curve estimator (Y-axis) vs. survival times (X-axis) of *T. granarium* until lay the first egg along with their 95% confidence bands at $30°C$, $35°C$ and $40°C$ respectively. Censored observations appear when *T. granarium* die before lay any egg and are symbolized by the "+" symbol. In the legend we report the median and its 95% C.I.

Table 2.4 Survival times (mean, std. error, median and 95% CI) until *T. granarium* lay the first egg within temperature levels.

Survival estimates until laying first egg		
Temperature	Mean $\pm$ se	Median and 95%CI
$30°C$	71.7 ± 1.8	73 (58, 76)
$35°C$	46.9 ± 1.7	45 (38, 49)
$40°C$	81.3 ± 3.4	- (68, -)

2.4.5.2 Parametric models

In order to assess a parametric fit to the *T. granarium* survival times, the Exponential, Weibull, Lognormal and Loglogistic distributions were considered.

The posterior means (95% Cr.I. s) for the Bernoulli parameter p are: 0.42 (0.28, 0.58), 0.50 (0.35, 0.64) and 0.96 (0.94, 0.98) at $30°C$, $35°C$ and $40°C$ respectively.

The AIC was estimated at 4096.40, 4040.70, 3772.10 and 3691.50 for the Exponential, Weibull, Lognormal and Loglogistic distributions respectively when time until death is consideredand 708.90, 20926.90, 570.10, 574.40 respectively in the case that time until first egg is studied (Table 2.5). It is evident that the smallest values are achieved by the Loglogistic model when the event is concerned with a *T. granarium* death, while the Lognormal model has the best fit when examining the time until a *T. granarium* lays the first egg (Fig. 2.6). The minimum percentage of zeros in the number

Table 2.5 Parametric survival times until death and until lay first egg (Dev, AIC and d.f.) for the Exponential, Weibull, Lognormal and Loglogistic models.

	Model	Dev	AIC	d.f.
Survival times until death	Exponential	-2046.2	4096.4	2
	Weibull	-2017.3	4040.7	3
	Lognormal	-1883.1	3772.1	3
	Loglogistic	-1842.8	3691.5	3
Survival times until lay first egg	Exponential	-352.4	708.9	2
	Weibull	-10460.4	20926.9	3
	Lognormal	-282.0	570.1	3
	Loglogistic	-282.4	574.4	3

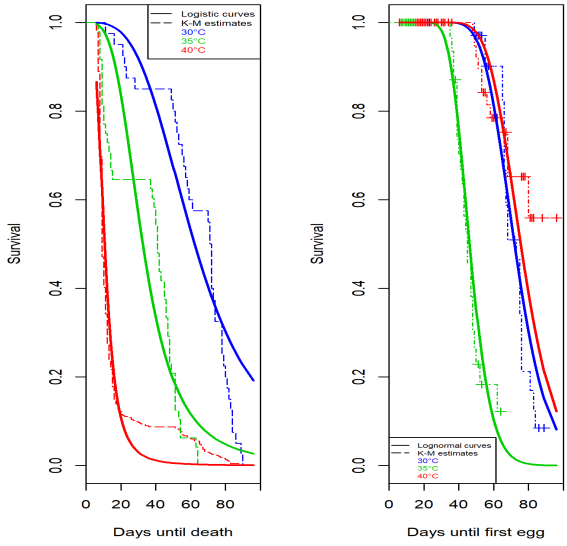

Figure 2.6 Logistic Mo1del and Lognormal Model probability (Y-axis) vs. survival times of *T. granarium*(X-axis) until death (left) and until laying the first egg (right) along with the Kaplan-Meier survival time estimates at $30°C$, $35°C$ and $40°C$ respectively.

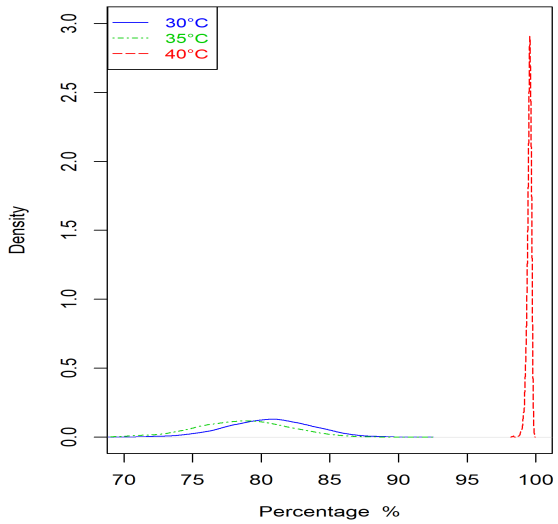

Figure 2.7 Plot of the posterior densities (Y-axis) for the probability of failing to lay eggs vs. the percentage of zeros (X-axis), at $30°C$, $35°C$ and $40°C$ respectively.

Table 2.6 Estimated results using the formula (2.30) of 1. probability of failing to lay eggs p(no egg)$= p + (1-p) \cdot e^{-\lambda}$, 2. probability p of excess zeros and 3. Poisson rate λ and their 95% Cr.I. in *T. granarium* data for each of the three temperature levels.

Temp	p(no egg)	95% Cr.I.	p	95% Cr.I.	λ	95% Cr.I.
$30°C$	0.805	(0.740, 0.863)	0.425	(0.281, 0.576)	0.431	(0.404, 0.459)
$35°C$	0.786	(0.716, 0.851)	0.498	(0.353, 0.638)	0.572	(0.533, 0.612)
$40°C$	0.996	(0.991, 0.998)	0.963	(0.941, 0.980)	0.143	(0.114, 0.161)

of *T. granarium* offspring for all temperature groups is over 0.40. Using the Bayesian paradigm to model the number of eggs, we can visualize the posterior densities of various ecological process components, such as the probability that the *T. granarium* beetle does not lay eggs, as depicted on Fig. 2.7. In addition as shown in (Table 2.6), the aforementioned probability is 0.81 (with a 95% Cr.I. of (0.74, 0.86)) at $30°C$, 0.79 (95% Cr.I.: (0.72, 0.85)) at $35°C$ and substantially higher at 0.996 (95% Cr.I.: (0.991, 0.998)) at $40°C$. The rate that *T. granarium* lay eggs in daily basis when they are active to reproduce is expressed by the lambda parameter of the Poisson distribution. Its posterior mean and 95% Cr.I. is 0.43 and (0.40, 0.46) at $30°C$, 0.57 and (0.53, 0.61) at $35°C$, 0.14 and (0.11, 0.16) at $40°C$ respectively. After performing the Wald test in the lambda parameters across the three temperature groups, we get "Bayesian p-values" which are less than 0.01, suggesting that there are significant differences in the number of eggs produced when comparing the three temperature groups, with the best performance observed at $35°C$ and the worst at $40°C$ respectively. Inspecting the standardised residuals suggests that no apparent pattern is emerging and no influential individual val-

ues stand out, indicating that the ZIP Model has good fit and explains reasonably well the randomness that stems from the inherent variability of *T. granarium* data.

2.5 Chapter summary

Our study revealed a number of findings on the biology of *T. granarium* and a comprehensive description of the survival and reproductive schedules of this invasive species in three distinct temperatures. We obtain further evidence on its growth rate, allowing for potential application in pest management. It was found that temperature seriously affects its population increase. The knowledge of the insect's potential growth rate also facilitates for estimation of its population through time, and therefore its potential outbreak. Adding uncertainty to the demographic cohort measures via bootstrap allows for hypothesis testing for different temperature levels. Specifically, at $40°C$ the value of the intrinsic rate of increase is negative, indicating that at this temperature the population tends to extinction, although *T. granarium* is considered a highly heat-tolerant species (Lindgren et al. 1955, Lindgren and Vincent 1959). At 30 and $35°C$ the positive values of the intrinsic rate of increase indicates that in this temperature range *T. granariun* is able to increase its population size, as well as its potential to spread, becoming more harmful in stored-products. The fact that there is a significant difference in the mean generation time between 30 and $35°C$ but not in the other demographic parameters may appear somewhat unexpected. The mean generation time represents the average time for a population to increase by a factor equal to the net reproductive rate. This result is biologically interpretable, since the net reproductive rate depends on cohort survival, which is lower for *T. granarium* at $35°C$. However, as the values of the intrinsic rate of increase and the doubling time did not differ significantly, we expect that the same applies for the insect's growth rate between these temperatures. According to the fit of the Briere model, the minimum and maximum temperatures for *T. granarium* population increase are roughly 18.44 and 40.00°C respectively. In this range of temperatures *T. granarium* is able to increase its population. This is important for the management of this species, considering its economic importance and further spreading in the world, as well as its mass-rearing, allowing efficient breeding in the insectary (Carey and Vargas 1985). The *T. granarium* intrinsic rate of increase shows an increasing trend until 34.52°C, where it reaches its maximum value. The subsequent decrease at higher temperatures is probably due to the determinental effect of these temperatures on its survival and reproductive capacity. According to the model's predictions, temperatures around 34°C are optimal for population growth of *T. granarium*, whereas temperatures in the area of $40°C$ lead to population decrease. These results clearly indicate that the population development of *T. granarium* is strongly affected by temperature. It should be noted that elevated temperature levels, which favor the popu-

lation increase of *T. granarium*, are responsible for potential outbreaks of this species, an issue that leads to considerable losses of the infested commodities (Kavallieratos et al. 2017b). Even when the initial population of *T. granarium* consists of a small number of larvae, it can increase fast under favorable temperature conditions and commodities (Kavallieratos et al. 2017b). Temperatures from $30°C$ to $35°C$ support the development of high numbers of *T. granarium* larvae, that is the most difficult life stage of this species to be controlled on stored commodities, especially on wheat (Kavallieratos et al. 2017b,a). Given that the efficacy of several insecticidal active ingredients against stored-product insect pest species varies among different levels of temperature (Kavallieratos et al. 2011, Athanassiou and Kavallieratos 2014, Boukouvala et al. 2017), potential optimization of chemical control measures should seriously take into account the combination of toxicants and temperature when applied against *T. granarium*. It is recommended to control this species when its numbers are still low as a way to moderate its population growth (Kavallieratos et al. 2017b). Also, based on our findings, since the population of *T. granarium* decreases at $40°C$, we could suggest a further rapid decline when insecticidal applications are targeted at the above temperature level on stored wheat. This is a realistic scenario, given that *T. granarium* is established in hot and dry environments (Banks 1977, Hill 2002, Lindgren et al. 1955). Our results clearly indicate that the population of *T. granarium* increases with temperature up until $34.52°C$. This is an important finding suggesting that global warming favors the increase of the population of this species. International trade in conjunction with global climatic change favors the dispersal of invasive species, like *T. granarium* (Kavallieratos et al. 2017b). Therefore, locations that are free of *T. granarium* but exhibit variable climatic conditions, compatible with those where *T. granarium* is already present, established or even intercepted should be on alert for the potential arrival of *T. granarium*. For example, the USA Government pays particular attention on phytosanitary measures and application of insecticides which aim to control *T. granarium* at the entry points of the country that are related to international trade (Hagstrum et al. 2012, Ghimire et al. 2017, Arthur et al. 2018). The fact that about 84% of *T. granarium* intercepted at the US ports between 1985 and 2010 were *T. granarium*, while after 2010 *T. granarium* interceptions have been dramatically increased in the USA and several countries of Northern and Southern Europe, reveals the potential risk of further and rapid expansion of this species worldwide (Hagstrum et al. 2012, Kavallieratos and Boukouvala 2018). The reproductive value of females, that is the contribution an individual of a particular age will make to future generations (Carey 1993), increases until a specific age. This is due to the early mortality of the pre-reproductive age classes of *T. granarium* and the subsequent increase of the age-specific fecundity. Thereafter, a decrease to the age-specific fecundity has a negative effect on the reproductive value which declines to zero for the older ages. Individuals of roughly 63, 42 and 21 days-old at 30, 35 and $40°C$ respec-

tively reach their maximum reproductive potential. The expected remaining lifetime decreases until a specific age at 35 and $40^{\circ}C$ due to early mortality, thereafter increases due to decreasing mortality, followed by an ultimate decrease. On the other hand, the expected remaining lifetime at 30ºC is characterized by a continuous decrease due to no remarkable early mortality. The process of *T. granarium* laying eggs was modeled by a Zero Inflated Poisson model (Diane 1992). Statistical learning for models of this kind represents a non-standard problem due to irregularities in the likelihood function and adopting a sampling-based approach to inference such as MCMC (Ghosh et al. 2006, Gelman et al. 2014) offers a substantial advantage, including the ability to estimate the complete posterior distribution of the Poisson rate and the probability of excess zeros. The separation of this probability at $40^{\circ}C$ compared to the other two temperatures is immediately apparent by simple visual inspection and this is represents a desirable feature of the proposed statistical analysis. The stochastic approach to demography offers a number of additional advantages. Here we present an effort towards the parametric characterisation of the different durations (Kalbfleisch and Prentice 2011) which represent the distinct components of the underlying biological process and in future research we shall endeavour to study the universality of these distributions by examining the parametric forms of these durations for related species. In addition, exploring the distribution of the time to the first birth naturally gives rise to an independent censoring mechanism necessitating a survival type of analysis for this component of our data. Investigating for influential individuals is of paramount importance for robust statistical results. Such considerations are relatively straightforward when adopting a stochastic approach to demography and this aspect was examined in the present study by leveraging upon the posterior density (Gelman et al. 2014), suggesting that the model appears to accommodate all the individual data reasonably well since no major departure from the bulk of the observations was observed. In summary, the use of a deterministic approach of *T. granarium* growth provides estimates of its reproductive potential, an issue that should be taken in account in the study of its biology and be considered as an important component in the design of pest's management strategies. Furthermore, our approach could be considered as an additional tool in a broader sense, combined with models related to international trade and climatic change, since these models alert specialists towards early detection strategies against invasive species and consequently their successful control (Kavallieratos et al. 2017b, Colunga-Garcia et al. 2013, Douma et al. 2016, Kriticos et al. 2013, Yemshanov et al. 2014). In addition, stochastic modelling of the variables (characteristics) of interest for *T. granarium* like their survival time, their time until first egg emerges or the number of eggs lying, provides an assessment of the variability for such variables, thus offering plausible ranges for use in alternative conditions (e.g. temperature, relative humidity, commodity), for comparison with different but related species. Also, the stochastic models of this study allowed for

checking model fit and the characterization of the most suitable distribution for each component of the system, allowing respectively for robust results and casting the durations involved in this particular species within a wider taxa.

Chapter 3

Inference for developmental rate in Ecology

3.1 Introduction

In this chapter, we suggest statistical models that are using distinct non-linear ecological models and data-sets with zeros to explain the developmental rates of arthropods. Furthermore, different computational methods such as HMC and ADVI are compared. The information criteria and marginal likelihood estimates are used to compare models. Furthermore, BMA is used to generate robust estimates for the parameters of interest. The current chapter's work was submitted for review to the journal of Environmental and Ecological Statistics in Summer 2021.

Studying the population evolution of arthropod pests, as well as of biological control agents, is of great importance for the crop primary production, agricultural infrastructures, spreading diseases and consequently the economy (Bradshaw et al. 2016). Temperature and body size are two major determinants that influence the metabolic, survival, growth and reproduction rates which control the ecological processes at all levels of arthropods' life (Brown et al. 2004). Biological control is facilitated when the climate responses of biocontrol agents are understood, especially to temperature. The thermal thresholds for insect development can be estimated using several functional forms (Kontodimas et al. 2004).

In all but the simplest cases, mathematical modelling is an indispensable tool for understanding the resulting developmental scheme (Kontodimas et al. 2004). However, fitting ecological models for developmental rates is not straightforward, typically because the mathematical forms are not linear (Papanikolaou et al. 2019) and the actual biochemical reactions of insects or environmental factors responsible for their growth may remain unobservable. Therefore, we adopt the Bayesian paradigm to population dynamics' modelling and inference since it naturally accounts for latent parameters and their uncertainties. Nonetheless, there are significant challenges in designing statistical methods that work efficiently in a wide range of ecological applications. STAN (Carpenter et al. 2017) provides a BUGS-like interface to model building and the ability to run it via different languages and operating systems.

3.1.1 Motivation for modelling developmental rate

Insects and mites, as ectotherms, regulate their body temperature according to the environment they live in (Norris and Kunz 2012). This affects the rate of metabolism, i.e., the biochemical reactions that allow the processes of production and release of energy, as well as the synthesis of necessary molecules that serve as structural or functional components (Neven 2000). In fact, temperature affects the functionality of enzymes, which act as catalysts for these biochemical reactions (van der Have 2008). Consequently, within a range of temperatures in which insects and mites develop and reproduce, various biological features are affected (Broufas and Koveos 2001, Huey and Berrigan 2001*a*, Broufas et al. 2007, Nedvěd 2009, Papanikolaou et al. 2013, 2014). Thus, their performance is indebted to several temporal fluctuations in terms of population size through time. The empirical finding of the initial increase in the growth rate of insects and mites in relation to temperature, followed by its sharp decline, formed the basis for the development of various mathematical models of its description (Kontodimas et al. 2004). These models allow the estimation of the lower and upper thermal limits, i.e. the lowest and highest temperature, respectively, at which the growth rate is zero, as well as the temperature at which it receives its maximum value. Understanding populations' growth rate is of importance, as their assessment can lead to decisions on their management (Hare et al. 2011), particularly under the pressure of climatic change (Bradshaw et al. 2016).

3.1.2 Historical overview with models and techniques used in literature

This thesis investigates some popular non-linear ecological models that describe the rate of insects' and mites' development within a Bayesian context. We explore the computational and statistical efficiency of HMC (Betancourt 2017, Neal 2010) and VBI (Blei et al. 2017), a challenging task in the present setting due to distinct features of the entertained models, including truncation. Both the widely used Gaussian distributional assumption and a newly-developed Inverse Gamma-based version are explored in order to model the developmental rate distribution of insects and mites. We compare the models using information criteria, marginal likelihood estimates and graphical tools. In addition, model averaging techniques are used to provide robust estimates of the parameters of interest. A distinct feature of these models is that zero count data make one parameter of interest indeterminable and model fit potentially misleading. We propose ways to overcome this indeterminacy by applying the Zero Inflated Inverse Gamma distribution while carefully connecting the probability of non-zero development to the predictor.

The remainder of this work is structured as follows. The next section contains the ecological models we develop. Section 3.3 the computation methods used are presented.

Section 3.5 investigates two separate real-life examples without and with zero-rates, and Section 3.6 reports our findings, while the chapter concludes with discussion.

3.2 Non-linear ecological models

A variety of different linear and non-linear equations have been used to describe the rate of development of insects and mites and to estimate their thermal limits. Such linear approximations enable the calculation of lower developmental threshold and thermal constant within a small temperature range, usually 15-30°C (Campbell et al. 1974, Wagner et al. 1991, Jarošík et al. 2002, Kontodimas et al. 2004). However, the relationship between development and temperature becomes non-linear outside that range. Thus, in order to accurately predict developmental rates across the spectrum, the use of non-linear ecological models is required (Wagner et al. 1991, Kontodimas et al. 2004, Damos and Savopoulou-Soultani 2012).

In the present work, four commonly used non-linear ecological models are considered. Specifically, the Bieri (Bieri et al. 1983), the Briere (Briere and Pracros 1998, Briere et al. 1998, 1999), the Analytis (Analytis 1981) and the Lactin (Lactin et al. 1995) models are implemented. Developmental time is the duration between life stages of the insects or mites (Wagner et al. 1991). The response variable $y(T;\theta)$ describes the developmental rate and it is defined as the reciprocal of the days until the completeness of a particular developmental event. Herein, T denotes the predictor variable, the absolute temperature measured in Celsius degrees, while θ denotes the parameter vector of the model. Typically, the expected developmental rate, $r(T;\theta)$ is modelled and the four aforementioned models are presented below.

3.2.1 Bieri Model

In the Bieri model, the developmental rate is defined as

$$r(T;\theta) = \alpha \cdot (T - T_{m_1}) - \beta^{(T-T_{m_2})} \tag{3.1}$$

where α, β, T_{m_1} and T_{m_2} are the model parameters. In particular, the values of T_{m_1} and T_{m_2} lie close to the real lower and upper thermal thresholds T_{min} and T_{max}, at which the development starts or ceases respectively. The exact values of the thermal thresholds are derived implicitly as the lower and upper roots respectively of the response function. The parameter α corresponds approximately to the rate of increase in the linear model at vital temperatures (Bieri et al. 1983). The response variable in (3.1) is a concave function of the temperature. Parameter α is defined in the interval $(0, 1)$ while β determines the decrease of the developmental rate at higher temperatures (Bieri et al. 1983) when β exceeds unity. According to (3.1), we observe that:

$$r(T_{m_1}) = -\beta^{(T_{m_1} - T_{m_2})} \text{ and } r(T_{m_2}) = \alpha(T_{m_2} - T_{m_1}) - 1,$$

which suggests that $r(T_{m_1}), r(T_{m_2}) \in (-1, 0)$ and leads to the following inequality

$$r(T_{m_1}) < r(T_{min}) < r(T_{max}) < r(T_{m_2}).$$

The temperature at which the maximum developmental rate occurs is called optimum and it is denoted by T_{opt}. In the Bieri model, it is given by

$$T_{opt} = T_{max} + \frac{\log \alpha - \log(\log \beta)}{\log \beta}.$$

In Fig. 3.1 some curves are generated by the Bieri model in (3.1) when $T_{m_1} = 5°\text{C}$ and $T_{m_2} = 35°\text{C}$ while the other parameters vary.

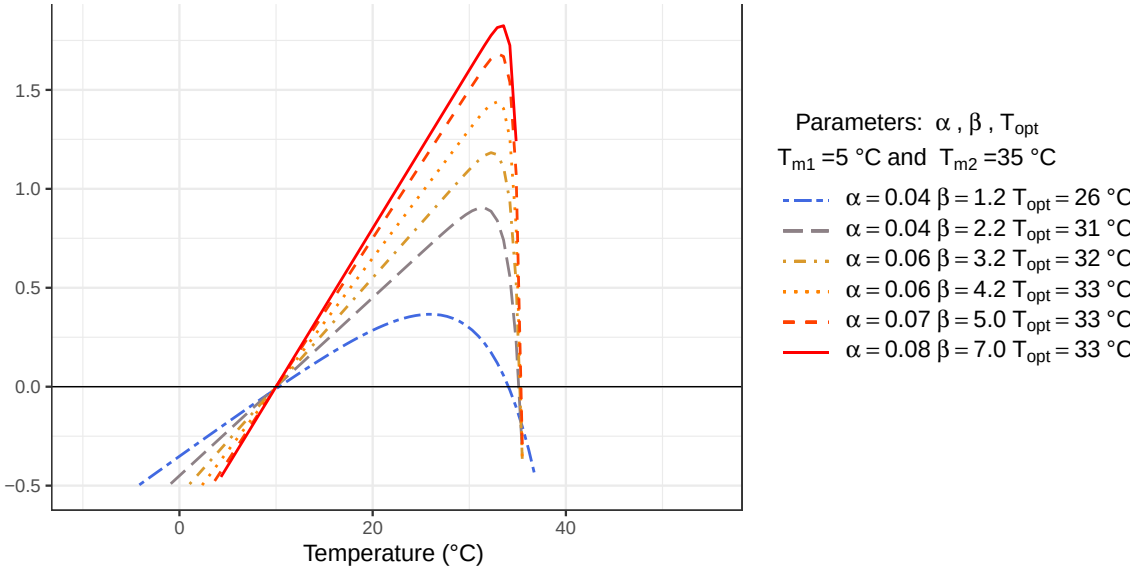

Figure 3.1 Bieri developmental rates.

3.2.2 Briere Model

The Briere model we already seen in section (2.2.2) is the most popular and parsimonious model. The developmental rate is defined as

$$r(T;\theta) = \begin{cases} \alpha \cdot T \cdot (T - T_{min}) \cdot \sqrt{(T_{max} - T)} & \text{for } T_{min} < T < T_{max} \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

where α, T_{min} and T_{max} are model parameters. Particularly, T_{min} and T_{max} are exactly the lower and upper thermal thresholds at which the development starts or ceases respectively while parameter α is an empirical constant (Briere et al. 1998). The response variable in (3.2) is again a concave function of the temperature. Parameter α is defined

in $(0, 1)$ whereas the existence of the square root in (3.2) ensures that the developmental rate declines sharply at higher temperatures. The optimum temperature in the Briere model is given by

$$T_{opt} = \frac{1}{10}\left\{4 \cdot T_{max} + 3 \cdot T_{min} + \sqrt{(4 \cdot T_{max} + 3 \cdot T_{min})^2 - 40 \cdot T_{min} \cdot T_{max}}\right\}.$$

In Fig. 3.2 some curves are created by the Briere model in (3.2) when $T_{min} = 5°C$ and $T_{max} = 35°C$ while the parameter α varies.

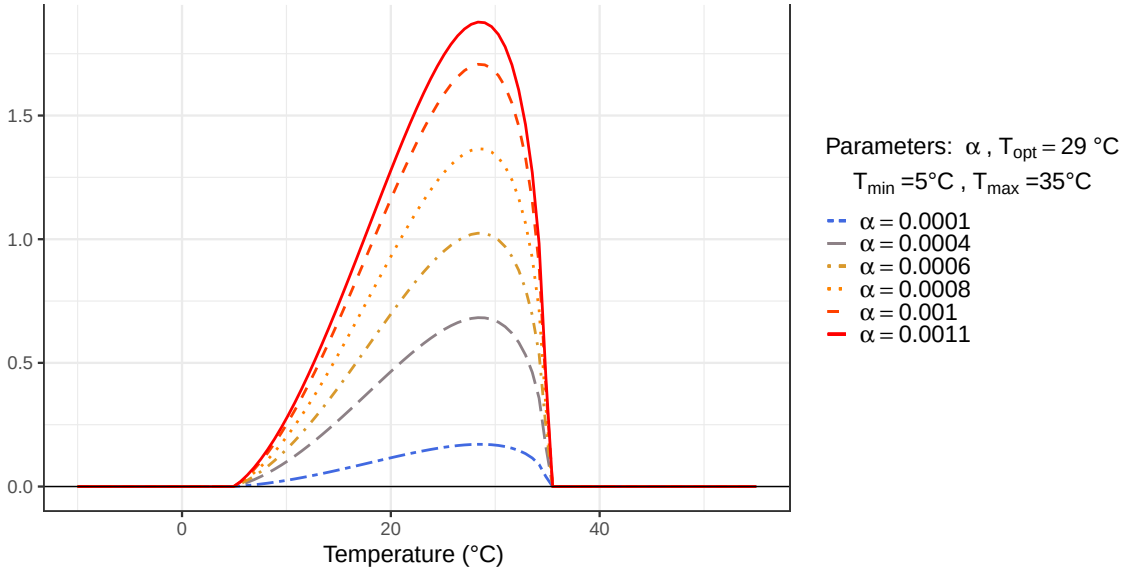

Figure 3.2 Briere developmental rates.

3.2.3 Analytis Model

The developmental rate in Analytis model is defined by

$$r(T;\theta) = \begin{cases} \alpha \cdot (T - T_{min})^n \cdot (T_{max} - T)^m & \text{for } T_{min} < T < T_{max} \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

where α, T_{min}, T_{max}, n and m are parameters of this model. The exponents n and m in (3.3) are empirical constants that determine the rate of growth and decrease of the developmental rate respectively (Analytis 1981). Both these parameter take values in $(0, +\infty)$ but in order to reduce the computational burden we may restrict them in a subset of the form $(0, c)$, for some constant $c > 0$. Finally, α takes values in $(0, 1)$ interval. The optimum temperature in this model is given by

$$T_{opt} = \frac{n \cdot T_{max} + m \cdot T_{min}}{n + m}.$$

The Analytis model has a multiplicative polynomial structure in which the exponents change as parameters to be estimated. Such a structure needs some empirical driven tuning when defining its thermal parameters space. Especially, in the case of one dataset, we assumed that T_{min} is greater than $4°C$. Some curves generated by the Analytis model in (3.3) are depicted in Fig. 3.3 when $T_{min} = 5°C$ and $T_{max} = 35°C$, while parameters n and m vary.

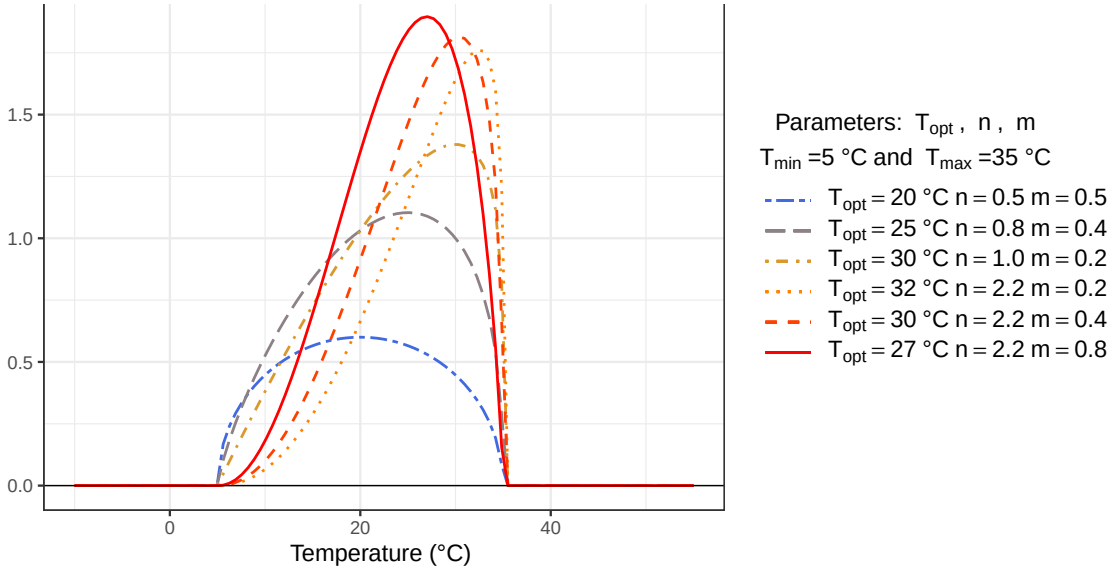

Figure 3.3 Analytis developmental rates.

3.2.4 Lactin Model

The Lactin model includes four parameters and the developmental rate is defined as

$$r(T;\theta) = \lambda + e^{\rho \cdot T} - e^{\rho \cdot T_m - \frac{(T_m - T)}{\Delta}} \tag{3.4}$$

where T_m is associated with the upper thermal threshold T_{max} since it tends to this value when λ tends to be zero. Parameter λ represents an asymptotic level of the developmental rate value in (3.4) that is approximated when predictor T tends either to $-\infty$ (extremely low temperatures) or to the threshold parameter T_m. Thus, in the event that the λ is non-negative, the Lactin ecological function does not have a lower thermal threshold (T_{min}) and at the same time $T_m \leq T_{max}$ as the developmental rate value in (3.4) is limited above zero level, at which the maximum thermal thresholds undergoes. In the case that λ is negative, $T_m > T_{max}$ and T_{min} sample space is in the $(-\infty, T_{max})$ interval. Parameter Δ is positive and it determines the descent steepness of the developmental rate. It expresses the temperature range between the value at which the response function begins to descend and the value of the T_m parameter. When Δ is less than one, the rate of descent is very high, although in the other case the rate of descent

is lower and similar to the other ecological models. This feature triggers discontinuity and lack of fit problems. Hence, in this work we define Δ in $(1, +\infty)$ interval in order to avoid such problems. Parameter ρ describes the acceleration of the function from low temperatures to the optimal temperature (Lactin et al. 1995). The response function in (3.4) has one inflection point, a maximum point at the optimum temperature and asymmetry about this point (left skewed). Also the function has a sharp drop after the optimum temperature, which is achieved by setting ρ in $(0, \Delta^{-1})$. The actual upper thermal threshold T_{max} is evaluated as the higher root of response function in (3.4). The optimum temperature in the Lactin model is given by

$$T_{opt} = T_m - \frac{\Delta \cdot \log(\rho \cdot \Delta)}{\rho \cdot \Delta - 1}. \tag{3.5}$$

In addition, the temperature at the inflection point of the Lactin curve is given by

$$T_{inf} = T_{opt} - \frac{\Delta \cdot \log(\rho \cdot \Delta)}{\rho \cdot \Delta - 1}. \tag{3.6}$$

In Fig. 3.4 some curves are created by the Lactin model in (3.4) when $T_m = 35°C$ and the parameters Δ, ρ and λ vary.

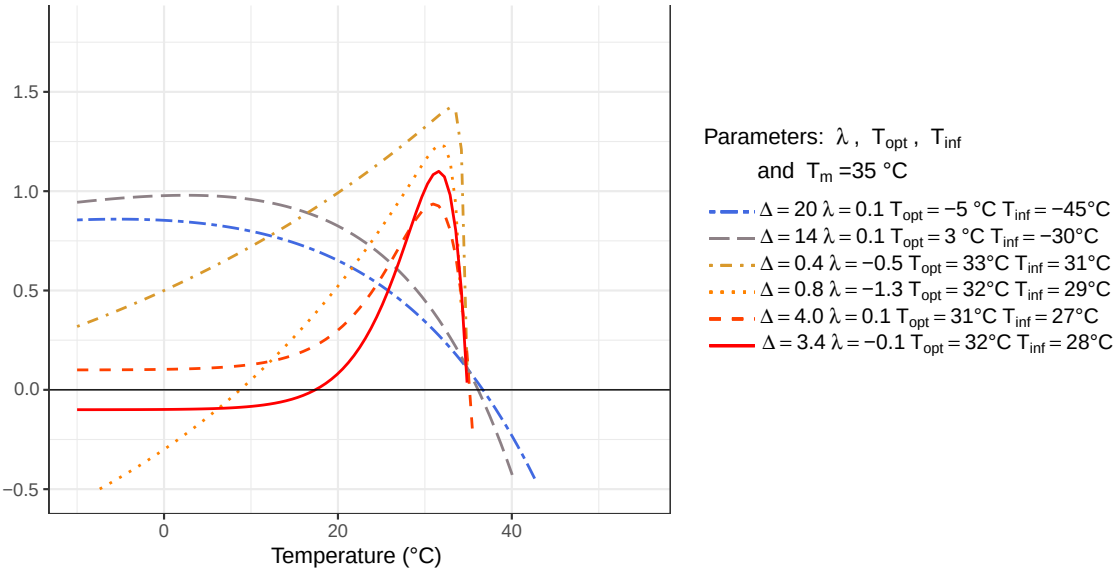

Figure 3.4 Lactin developmental rates.

3.2.5 Ecological features of the models

There are some basic common features in all of the above-mentioned ecological models.

There is no growth below the lower temperature threshold T_{min} or above the upper temperature threshold T_{max}. Specifically, in the case of the Briere and Analytis models,

the developmental rate is positive and is defined only between the two parameters of the thermal thresholds. Also, the developmental rate in the Bieri and Lactin models can take negative values that cannot be interpreted. To accommodate these characteristics, we use initial values of the parameters that allow the ecological function to receive positive values.

The developmental rate is an asymmetric curve left skewed of its maximum point. It increases and reaches a maximum at optimal temperature while it declines rapidly down to zero at the higher temperature threshold T_{max} that is considered as lethal temperature. It includes an inflection point, with the exception of the Bieri model which is a concave function of the temperature. The structure of the Lactin model makes it susceptible to a type of exponential pattern in data-values (Fig. 3.4). The Briere model, on the contrary, has a specific structure, which has the first derivative of class $\mathcal{O}\left(T^{2.5}\right)$ and can hardly trace exponential changes in data values (Fig. 3.2). Also, the model of Analytis has polyonimic structure as the model of Briere but it does contain more degrees of freedom because the exponents it includes are unknown parameters (Fig. 3.3). The Bieri model also adopts an exponential reduction after the thermal optimum threshold but can only follow the developmental increase of the dataset linearly (Fig. 3.1). All four ecological models have been used in the literature to provide reasonable estimates of the thermal thresholds of several anthropods' developmental rates at various stages (Bieri et al. 1983, Kontodimas et al. 2004, Aghdam et al. 2011). However, the Bieri model is underutilized in the literature, so we include it in this study to gain a better understanding of ecological models that describe temperature-dependent development. Additionally, in (Kontodimas et al. 2004, Aghdam et al. 2011) the ecological models were compared based on the accuracy of the real data thermal threshold estimates, the adjusted coefficients of regression (R^2), and the residual sum of squares values. In summary, the Briere and Analytis models, appear to overestimate and underestimate the upper and lower thermal thresholds, respectively, whereas the Bieri and Lactin models appear to meet the majority of the criteria used in the comparison. Despite these minor differences, all of the above models appear to provide higher R^2 values than other models in real-data applications in the literature and we include all four in the current work.

3.3 Measurement error

Probabilistic random error due to chance and systematic error due to data with excessive zeros is added in order to include uncertainty in the ecological models already listed. In section 3.3.1, we present the notation adopted and the probability schemes implemented in this analysis.

3.3.1 The likelihood of the data

Let y_i represent the observed developmental rates of the i^{th} individual observed at T_i temperature, where $i = 1, \ldots, N$. We consider y_i to be independent response variables counted as the reciprocal of the number of days until the development of the i^{th} individual takes place and the range of its value lies in the $(0, 1)$ interval. Furthermore, let y, T and θ be the vectors of the response, the predictor and the parameters respectively then the conditional expectation $E(y|T, \theta)$ is considered to vary according to the ecological function $r(T; \theta)$ presented in (3.1), (3.2), (3.3) and (3.4) respectively.
The data distribution is denoted by

$$p(y|T, \theta) = \prod_{i=1}^{N} p(y_i|T_i, \theta).$$

In this study we consider the Gaussian and the Inverse-Gamma distributions as the distribution of the response data y_i. Specifically, the Gaussian distribution is used broadly in the literature as a good approximation to most unimodal distributions with finite variance due to the general form of the Central Limit Theorem. Thus, it can be used as to approximate a more complicated model likelihood of the data. The non-linear model of the independent response rates has the Gaussian distribution given by

$$y_i \sim N(r(T_i; \theta), \sigma^2) \tag{3.7}$$

where the mean of the Gaussian likelihood is driven by the respective ecological model, whereas its standard deviation σ is considered as an unknown parameter. We consider a weakly informative prior distribution for σ like: the Inverse-Gamma Inv$\Gamma\left(10^{-3}, 10^{-3}\right)$. The non-linear model used in (3.7) is a constant variance model among different temperatures. In addition, it allows for zero observed rate values, which occur when the stage of the insect does not change in perpetuity. Such modelling, however, has the disadvantage of lack of interpretability in the case of estimated negative rate values.

On the other hand, the Inverse-Gamma distribution is a plausible alternative for modelling positive observed rates, as it handles positive values that describe ratios such as developmental rates whose inverse are positive counts (such as days passed until the expected development occurs) that can be described by the Gamma distribution. The non-linear model of the independent response rates has the Inverse-Gamma distribution given by

$$y_i \sim \text{Inv}\Gamma(\zeta, (\zeta - 1) \cdot r(T_i; \theta)) \tag{3.8}$$

where ζ is the shape parameter of the Inverse-Gamma distribution. The mean of the distribution equals $r(T_i; \theta)$, whereas the variance equals $\frac{r(T_i;\theta)^2}{(\zeta-2)}$. The Inverse-Gamma likelihood in (3.8) is a natural alternative to model observed positive rates. Herein, its

mean is driven by the respective ecological function, whereas its variance depends both on the shape parameter ζ and the ecological function $r(T_i;\theta)$ as well, which allows the variance to be temperature dependent. For the prior distribution of the shape parameter ζ we have chosen a weakly informative $\Gamma\left(10^{-1}, 10^{-2}\right)$.

3.3.2 Zero-rates case

There are situations in which any insect development does not occur throughout the cohort study. This is indicated by zero values in the response variable, which can theoretically be interpreted as the number of days required for the insect to move to its next stage never ending, implying that the developmental rate is the reciprocal of infinity. In such cases, the MCMC sampling procedure can be extended either to include prior information about the case of no development by adjusting the prior knowledge of the parameters concerned or to include a zero-inflation scheme. For this work, we suggest the use of a Zero Inflated Inverse Gamma distribution, which gives zero value with probability p_i for the i^{th} insect observed at T_i temperature. Particularly, the probability density function of the observation y_i is:

$$\begin{aligned} P(y_i|T_i,\theta) &= \begin{cases} p_i & \text{if } y_i = 0 \\ P_{\text{Inv}\Gamma}\left\{\zeta, (\zeta-1)\cdot r(T_i;\theta)\right\}\cdot(1-p_i) & \text{if } y_i \neq 0 \end{cases} \\ \text{logit}\,(1-p_i) &= \left[c\cdot\left\{r(T_i;\theta)-k\right\}\right] \Rightarrow p_i = \frac{1}{e^{\left\{c\cdot(r(T_i;\theta)-k)\right\}}+1} \end{aligned} \tag{3.9}$$

where c is a constant positive parameter that has the opposite order of magnitude of the sample mean $\overline{y}$, while k is the inflation point of logit link function where the probability of zero p_i is equal to $\frac{1}{2}$. The constant k is associated with the constant c and can be chosen so that the zero rate of development matches the probability p_i at a predetermined level like 0.9. According to the real data example presented below in the results section, the proposed values that satisfy the above criteria for constants c and k are 10^2 and $5\cdot10^{-3}$, respectively.

As the developmental rate of $r(T_i)$ increases to its maximum, the probability p_i in (3.9) decreases towards zero. On the other hand, when $r(T_i)$ tends to be a very small number, the probability p_i in (3.9) tends to be one.

3.3.3 Priors

In the Bieri model (3.1) the prior distribution of the parameters α, β, T_{m_1} and T_{m_2} are shown in Table 3.1 respectively.

In the Briere model (3.2), the transformation $\widetilde{a} = -\log(\alpha)$ is considered instead of the original parameter α. Weakly informative priors are considered for $\widetilde{a}$, T_{min} and T_{max} as shown in Table 3.1 respectively. In the Analytis model (3.3) the transformation $\widetilde{a} =$

Table 3.1 Priors of the parameters of the four ecological models.

Models	Parameters	Priors
Bieri	α	U(0,1)
	β	$\Gamma\left(2 \cdot 10^{-1}, 10^{-1}\right)$
	T_{m_1}	$\Gamma\left(10^{-1}, 10^{-2}\right)$
	T_{m_2}	$\Gamma\left(10^{-1}, 10^{-2}\right)$
Lactin	$^{\dagger}l = -\lambda$	$\Gamma\left(10^{-1}, 10^{-1}\right)$
	$^{\dagger}del = \frac{1}{\Delta}$	U(0,1)
	$^{\dagger}a = e^{(\rho - del) \cdot T_m}$	$\Gamma\left(10^{-2}, 10^{-3}\right)$
	ρ	U(0,1)
Briere	$^{\dagger}\widetilde{a} = -\log(\alpha)$	$\Gamma\left(10^{-1}, 10^{-2}\right)$
	T_{min}	$\Gamma\left(10^{-2}, 10^{-2}\right)$
	T_{max}	$\Gamma\left(10^{-2}, 10^{-3}\right)$
Analytis	$^{\dagger}\widetilde{a} = -\log(\alpha)$	$\Gamma\left(10^{-1}, 10^{-2}\right)$
	m	$\Gamma\left(10^{-1}, 10^{-1}\right)$
	m	$\Gamma\left(10^{-1}, 10^{-1}\right)$
	T_{min}	$\Gamma\left(10^{-2}, 10^{-2}\right)$
	T_{max}	$\Gamma\left(10^{-2}, 10^{-3}\right)$

† transformed parameter used.

$-\log(\alpha)$ is considered instead of α with the weakly prior distribution $\Gamma\left(10^{-1}, 10^{-2}\right)$. Additionally, the prior distributions utilized for parameters m, n, T_{min} and T_{max} are given in Table 3.1 respectively.

In the Lactin model (3.4) a new parametrization is used. Hence, the new transformed parameters are:

$$l = -\lambda \quad , \quad del = \frac{1}{\Delta} \quad , \quad a = e^{(\rho - del) \cdot T_m}$$

which are considered instead of the original parameters λ, Δ and T_m, respectively. By their definition, the new parameters are taking values in $\lambda \in \mathcal{R}$, $0 < del < 1$ and $a \in (0, \rho \cdot \Delta)$. The latter is derived by the definition of the optimum temperature in the Lactin model (3.5) and the fact that $T_{opt} > 0$. In addition, should the temperature at the inflection point T_{inf} be a positive number, then from (3.6) we can get that $a \in (0, \rho^2 \cdot \Delta^2)$. The prior distributions considered are shown in Table 3.1 respectively for each ecological model.

3.4 Bayesian inference

In the current section, we provide details of statistical model specifications in the Bayesian framework. Specifically, we provide a brief description of the most important features of Stan's implementation of HMC and VBI so the reader can get familiar with the tools that Stan is based on. We then provide model selection and model averaging techniques in order to compare the different ecological models and to explore and interpret the

parameters of interest combining predictions from all the four of them.

3.4.1 HMC and VBI techniques

The HMC method is a Monte Carlo technique that uses Hamiltonian dynamics in order not only to explore efficiently the target distribution but also to propose distant samples in the parameter space that do not exclusively depend on the current state of the Markov chain like considered in previous MCMC methodology (Neal 2010). In this way, many performance challenges are tackled like either the slow convergence due to the fact that the parameter space with high posterior support is not reached or the poor exploration of the target distribution due to its multi-modality or its shape irregularities. The existence of Hamiltonian dynamics in the system of the joint density mass function allows the preservation of volume and hence adequate trajectories can be used to define complex mappings of the parameter state space without the need to account for cumbersome Jacobian calculations (Barber et al. 2003). Thus, by carefully designing automated trajectory realizations in the Hamiltonian dynamics system, the Stan team managed to create an augmented software called STAN (Carpenter et al. 2017), which materializes HMC sampling for the parameters of interest.

Moreover, independently, Automatic Differentiation Variational Inference (ADVI) technique is referred to the machine learning field (Blei et al. 2017). The latter is a VBI method and posterior target distributions are approximated by choosing the closest distribution to a parametric family of tractable distributions like the exponential family via optimization. In order to achieve this point-wise estimations of the parameters of the family distribution are estimated so that the Kullback–Leibler 'KL' divergence function is minimized. Specifically, since the KL divergence is intractable the Evidence Lower Bound is maximized instead (Blei et al. 2017).

3.4.2 Model selection and model averaging

In case there are m models $(\mathcal{M}_1, \ldots, \mathcal{M}_m)$ under consideration, the posterior probability of the suitability of the i^{th} model given the data y, is given by

$$p(\mathcal{M}_i|y) = \frac{p(y|\mathcal{M}_i)p(\mathcal{M}_i)}{\sum_{k=1}^{m} p(y|\mathcal{M}_k)p(\mathcal{M}_k)} \tag{3.10}$$

where $p(\mathcal{M}_i)$ expresses the prior belief for the i^{th} model, while the $p(y|\mathcal{M}_i)$ is the model evidence also called 'marginal likelihood' and it can be interpreted as the likelihood over the space of models, marginalizing out the parameters of the i^{th} model. The ratio of the marginal likelihoods between two models $\frac{p(y|\mathcal{M}_i)}{p(y|\mathcal{M}_j)}$ is called Bayes factor and is the posterior odds of the null hypothesis that the i_{th} model fits better the data than the j_{th} model does when the prior probability of the null is one-half (Kass and Raftery

1995). The Bayes factor is used in order to give evidence for the most probable model given the data, when comparing two alternative models (Kass and Raftery 1995).

In the case of Bayesian model averaging, the model selection uncertainty is taking into account in statistical inference. The joint posterior $p(\mathcal{M}_i, \theta_i|y)$ of the i^{th} model with vector of parameters θ_i, using the Baye's rule, is proportional to the product of the likelihood of the i^{th} model times the prior distribution of the parameters $p(\theta)$ times the prior distribution $p(\mathcal{M}_i)$ (that expresses our uncertainty of the i^{th} model)

$$p(\mathcal{M}_i, \theta_i|y) \propto p(y|\theta_i, \mathcal{M}_i) \cdot p(\theta) \cdot p(\mathcal{M}_i). \tag{3.11}$$

The uncertainty of the i^{th} model given the data can then be re-expressed via the posterior probability $p(\mathcal{M}_i|y)$ defined by the ratio in (3.10), in case of existence of multiple models. The posteriors of the models can be thought as weights that are critical to the Bayesian model averaging as they can be used to extract useful weighted statistics from the data distribution while at the same time taking into account model uncertainties. Estimation of model parameters and model uncertainties can be achieved either by directly sampling from the joint posterior (3.11) or by approximating the marginal likelihood of each model independently and, accordingly, by controlling the outcomes with a view to formulating proper weights and proceeding with the calculation of the averaged statistics. For the former case, techniques like the reversible jump MCMC (Green 1995, George and McCulloch 1997) and variable selection samples (Carlin and Chib 1995, Kuo and Mallick 1998, Dellaportas et al. 2000) are used. On the other hand, for the later case, techniques of marginal likelihood approximations via thermodynamic integration (Friel and Pettitt 2008), bridge sampling (Meng and Wong 1996), importance sampling (Perrakis et al. 2014) or via information- criteria perspective like in (Kass and Raftery 1995) are used.

3.4.2.1 Information criteria

The criteria used in the current work are the Akaike information criterion 'AIC', the Bayesian information criterion 'BIC', the Deviance information criterion 'DIC', the Watanabe–Akaike information criterion 'WAIC' and the Leave-one-out cross-validation criterion 'Loocv'. Briefly, these criteria provide an approximation of the expected log predictive density for new-coming data while correcting bias from data usage. In particular AIC (Akaike 1974) is defined as the difference

$$AIC(\mathcal{M}_i) = -2\log(y|\widehat{\theta}_i) + 2k_i$$

where $\widehat{\theta}_i$ is the Maximum Likelihood estimate 'MLE' of the k_i parameters of the i^{th} model. Similarly, BIC (Schwarz et al. 1978) is defined as the difference

$$BIC(\mathcal{M}_i) = -2\log(y|\widehat{\theta}_i) + k_i \cdot \log(n)$$

where n is the sample size. In addition, DIC (Spiegelhalter et al. 2002) is defined as the following difference:

$$DIC(\mathcal{M}_i) = -2\log p(y|\widehat{\theta}_i) + 2p_{DIC}$$

where $\widehat{\theta}_i$ is the posterior mean of the parameters of the i^{th} model, whereas p_{DIC} is the effective number of parameters and it is evaluated following (Spiegelhalter et al. 2002, Gelman et al. 2014) by either

$$p_{DIC_1} = E_{\theta|y}\{-2\log p\left(y|\theta\right)\} + 2\log\left\{p\left(y|\widehat{\theta}_i\right)\right\},$$

or

$$p_{DIC_2} = \frac{Var_{\theta|y}\left\{\log p(y|\theta)\right\}}{2},$$

where $E_{\theta|y}\left\{\log p(y|\theta)\right\}$ is an expectation over the posterior density of θ, whereas $Var_{\theta|y}\left\{\log p(y|\widehat{\theta})\right\}$ is the variance of the log posterior density of the observed data y, over the posterior density of θ. Furthermore, WAIC (Watanabe 2010) is defined as the following difference:

$$WAIC(\mathcal{M}_i) = -2\sum_{j=1}^{N}\log E_{\theta_i}\left\{p(y_j|\theta_i)\right\} + 2p_{WAIC}$$

where $E_{\theta_i}\left\{p(y_j|\theta_i)\right\}$ is the expectation of the probability at y_j data point over the posterior distribution of the parameters of the i^{th} model, whereas p_{WAIC} is the effective number of parameters and it is evaluated following (Gelman et al. 2014) by either

$$p_{WAIC_1} = 2\sum_{j=1}^{N}\left[\log E_{p(\theta|y)}\left\{p(y_j|\theta)\right\} - E_{p(\theta|y)}\left\{\log p(y_j|\theta)\right\}\right]$$

or

$$p_{WAIC_2} = \sum_{j=1}^{N} var_{p(\theta|y)}\left\{\log p(y_j|\theta)\right\}$$

where $E_{p(\theta|y)}\left\{\log p(y_j|\theta)\right\}$ is the expectation over the logarithm of the posterior density of θ at y_j data point, whereas $var_{p(\theta|y)}\left\{\log p(y_j|\theta)\right\}$ is the variance of the log posterior density of the observed data y_j, over the posterior density of θ.

Furthermore, LooCV (Gelman et al. 2014) is defined as the following difference:

$$LooCV(\mathcal{M}_i) = -2\sum_{j=1}^{N} \log E_{\theta_i^{-j}} \left\{ p(y_j|\theta_i^{-j}) \right\} - 2\beta_{LooCV}$$

where $E_{\theta_i^{-j}} \left\{ p(y_j|\theta_i^{-j}) \right\}$ is the expectation of the probability at y_j data point over the posterior distribution of the parameters of the i^{th} model. The posterior distribution $p(\theta_i^{-j}|y_{-j})$ is sampled considering a partition of the data, leaving one data value (y_j) out of the original sample. The β_{LooCV} is a bias correction of the measure and it is evaluated following (Gelman et al. 2014) by

$$\beta_{Loocv} = \sum_{j=1}^{N} \log E_{\theta_i} \left\{ p(y_j|\theta_i) \right\} - \frac{1}{N} \sum_{\kappa=1}^{N} \sum_{j=1}^{N} \log E_{\theta_i^{-\kappa}} \left\{ p(y_j|\theta_i^{-\kappa}) \right\}$$

where $E_{\theta_i^{-\kappa}} \left\{ p(y_j|\theta_i^{-\kappa}) \right\}$ is the expectation of the probability at y_j data point over the posterior distribution of the parameters of the i^{th} model leaving out the κ^{th} observation.

3.4.2.2 Marginal likelihood estimation techniques

The marginal likelihood can be viewed as a normalizing constant $z_i = p(y|\mathcal{M}_i)$ of the density $q(\theta_i|y) = p(y|\theta_i) \cdot p(\theta_i)$ within the i^{th} ecological model that includes parameters θ_i. In the general scheme of comparing the two densities q_0 and q_1 of interest, as in the case of the Bayes factor of two models or in the case model's prior and posterior, a general path from q_0 to q_1 can be created according to (Gelman and Meng 1998) using a class of densities $p(\theta_i|y,t)$ on the same space indexed by the continuous auxiliary variable say $t \in [0,1]$. A key formula that links the corresponding normalizing constant $z(t)$ and the unormalized density $q(\theta_i|y,t)$ that correspond to the sampling distribution $p(\theta_i|y,t)$ is given by:

$$\frac{d}{dt} \log z(t) = \int \frac{1}{z(t)} \frac{d}{dt} q(\theta_i|y,t) p(\theta_i|t) d\theta_i = E_t \left\{ \frac{d}{dt} \log q(\theta_i|y,t) \right\}, \tag{3.12}$$

where the expectation is with respect to the sampling distribution $p(\theta_i|y,t)$.

In addition, another key formula of estimating a ratio of normalizing constants has been of great interest such as in computing likelihood ratios in hypothesis testing or in computational physics in estimating free energy differences, or in computing the Bayes factor in Bayesian framework (Meng and Wong 1996). The general formula is as follows:

$$\frac{z_1}{z_0} = \frac{p(y|\theta, t=1)}{p(y|\theta, t=0)} = \frac{E_0 \left\{ h(\theta) \cdot q(\theta|y, t=1) \right\}}{E_1 \left\{ h(\theta) \cdot q(\theta|y, t=0) \right\}} \tag{3.13}$$

where E_0 and E_1 expectations are with respect to posterior distribution densities $p(\theta|y, t = 0)$ and $p(\theta|y, t = 1)$ respectively, whereas the bridge function $h(\theta)$ is defined and over-

lapped by the common support of the former densities.

Using general formulas (3.12) and (3.13), several marginal probability evaluation schemes of the i^{th} model are derived (Gelman and Meng 1998). The power posterior sampling (Friel and Pettitt 2008), the importance sampling (Perrakis et al. 2014) and the bridge sampling (Meng and Wong 1996, Overstall and Forster 2010) techniques are used for the current work.

In the power posterior case, formula (3.12) is integrated with respect to variable t and $q(\theta_i|y,t)$ is substituted with density $p(\theta_i|y)^t p(\theta_i)$. The marginal likelihood $z_i = p(y|\mathcal{M}_i)$ is derived from logarithmic scale by the equation:

$$\log\{p(y|\mathcal{M}_i)\} = \int_0^1 E_{\theta_i|y,t}\{\log p(y|\theta_i)\}\, dt \tag{3.14}$$

where expectation $E_{\theta_i|y,t}$ is taken with respect to the density $p(\theta_i|y)^t p(\theta_i)$ which is defined as the power posterior at temperature t (Friel and Pettitt 2008).

Additionally, the standard error se_i for the ith model estimator (3.14), as shown in section B.1 of the appendix B is approximated by:

$$\hat{se}_i = \sqrt{\frac{(t_2-t_1)^2}{2}s_1^2 + \sum_{k=2}^{n-1}\frac{(t_k-t_{k-1})^2}{2}s_k^2 + \frac{(t_n-t_{n-1})^2}{2}s_n^2},$$

where t_k is the time after discretization $0 = t_0 < t_1 < t_k < t_n = 1$ and s_k is the standard error of the corresponding estimation $\log\{p(y|\mathcal{M}_i)\}$ given in (3.14).

In the case of importance sampling, the marginal likelihood is assessed by introducing the proper density function *g*. After sampling from the proposed density function *g*, the marginal likelihood is calculated as with respect to *g* as:

$$p(y|\mathcal{M}_i) = E_g\left\{\frac{q(\theta_i|y)}{g(\theta_i)}\right\}$$

Following (Perrakis et al. 2014), we use the density $q(\theta_i|y)$ equal to $p(y|\theta_i,\phi_i)\cdot p(\theta_i,\phi_i)$ and the auxiliary importance function *g* used is as follows:

$$g(\theta_i) = g(\theta_i,\phi_i) = p(\theta_i|y)p(\phi_i|y), \tag{3.15}$$

where (θ_i,ϕ_i) are the parameters of the i^{th} model divided into two blocks θ_i and ϕ_i which may or may not be independent. The right hand side of (3.15) is the product of the marginal posterior distributions of the block. Thus, the marginal probability which gives the target value is given as follows:

$$p(y|\mathcal{M}_i) = \iint \frac{p(y,\theta_i,\phi_i)}{g(\theta_i,\phi_i)} g(\theta_i,\phi_i)d(\theta_i,\phi_i) = E_g\left\{\frac{p(y,\theta_i,\phi_i)}{g(\theta_i,\phi_i)}\right\} \tag{3.16}$$

The standard error se_i of (3.16) as shown in section B.2 of the appendix B is:

$$\hat{se}_i = \sqrt{\frac{1}{K}\sum_{j=1}^{K}\left\{\frac{p(y|\theta^j,\phi^j)\cdot p(\theta^j,\phi^j)}{g\left(\theta^j\right)} - \hat{z}_i\right\}^2},$$

where $\widehat{z}_i$ is the estimation of the corresponding marginal probability (in the same form of (3.16)), while (θ^j, ϕ^j) are draws $j = 1, 2, \ldots, K$ from the importance function in (3.15).

Additionally, using an alternative version of (3.13) in (Meng and Wong 1996, Frühwirth-Schnatter 2004, Overstall and Forster 2010) the marginal likelihood of a single model is evaluated using bridge sampling by the formula:

$$z_i = p(y|\mathcal{M}_i) = \frac{E_g\left\{h(\theta_i)\cdot q(\theta_i|y)\right\}}{E_p\left\{h(\theta_i)\cdot g(\theta_i)\right\}}, \tag{3.17}$$

where E_g and E_p are the expectations with respect to $g(\theta_i)$ a so-called proposal distribution and to $p(\theta_i|y)$ the i^{th} model posterior distribution respectively.

The bridge function $h(\theta_i)$ is selected to minimize the relative mean-squared error of (3.13). Following (Meng and Wong 1996) the bridge function is specified by:

$$h(\theta_i) = \mathcal{C}\cdot\frac{1}{s_1\cdot q(\theta_i|y) + s_2\cdot p(y)\cdot g(\theta_i)}, \tag{3.18}$$

where $s_1 = \frac{N_1}{N_1+N_2}$, $s_2 = \frac{N_2}{N_1+N_2}$ and $\mathcal{C}$ is a constant. N_1 is the sample size from the posterior and N_2 is the sample size from $g(\theta_i)$.

The optimal bridge function in (3.18) includes the marginal likelihood under-assessment so that it cannot be evaluated directly. For this purpose the iterative method suggested by (Meng and Wong 1996) and applied in (Gronau et al. 2020) in R software (Team 2021) is used. The alternatives used in place of distribution g is either a multivariate normal distribution with mean vector and covariance matrix that match the respective posterior samples quantities or a standard multivariate normal distribution in combination with a warped posterior distribution of which the first three moments correspond to (Gronau et al. 2020).

Moreover, following (Frühwirth-Schnatter 2004) the relative mean square error $RE_i^2 = \frac{E\{\widehat{z}_i - z_i\}^2}{z_i^2}$ of (3.17) is evaluated by the formula:

$$\widehat{RE}_i^2 = \frac{1}{N_2}\frac{V_g\left\{f_1(\theta_i)\right\}}{E_g^2\left\{f_1(\theta_i)\right\}} + \frac{\rho_{f_2}(0)}{N_1}\frac{V_p\left\{f_2(\theta_i)\right\}}{E_p^2\left\{f_2(\theta_i)\right\}}, \tag{3.19}$$

where $f_1(\theta_i) = \frac{q(\theta_i|y)}{s_1\cdot q(\theta_i|y)+s_2\cdot g(\theta_i)}$, $f_2(\theta_i) = \frac{g(\theta_i)}{s_1\cdot q(\theta_i|y)+s_2\cdot g(\theta_i)}$,
$V_g(f_1(\theta_i)) = \int\left\{f_1(\theta_i) - E(f_1(\theta_i))\right\}^2 g(\theta_i)d\theta$ is the variance of $f_1(\theta_i)$ with respect to

$g(\theta_i)$. The term $\rho_{f_2}(0)$ in (3.19) corresponds to the normalized spectral density of the auto-correlated process $f_2(\theta_i)$ at the frequency 0.

Following (Gronau et al. 2020) the square root of RE^2 can be interpreted as coefficient of variation provided that the bridge sampling estimator $\widehat{z}_i$ is unbiased. Then the stantard error se_i of the bridge estimator is evaluated by the product $\widehat{se}_i = \widehat{RE} \cdot E(\widehat{z}_i)$

3.4.2.3 BMA weights

We can derive a weighted prediction $\tilde{y}$ over the m different models $\mathcal{M}_1, \mathcal{M}_2, \dots \mathcal{M}_m$ predictions $\hat{y}_1, \hat{y}_2, \dots \hat{y}_m$ by imposing appropriate weights $w_1, w_2, \dots w_m$.

$$\tilde{y} = \sum_{i=1}^{m} \hat{y}_i \cdot w_i \quad \text{and} \quad \sum_{i=1}^{m} w_i = 1.$$

In the Bayesian framework, model weights definition is straightforward. The model weights used are the posterior model weights $w_i = p(\mathcal{M}_i|y)$ given in (3.10) that represent the relative probability of each model given the data. So a major challenge is to estimate these Bayesian weights. Except for using the marginal likelihood estimations mentioned in previous section, we also investigate approximations of the weights by using the BIC for each model. In particular, model weights can be estimated through the following equations (Kass and Raftery 1995, Buckland et al. 1997):

$$w_i = \frac{e^{-0.5 \cdot (BIC(\mathcal{M}_i))}}{\sum_{j=1}^{m} e^{-0.5 \cdot (BIC(\mathcal{M}_j))}}. \tag{3.20}$$

Instead of BIC, the AIC, DIC, WAIC and LooCV are also used in (3.20). We investigate both approaches in insect observed rates and compare the results taking into account model complexity, data scarcity and Biological interpretation.

3.5 Real-life applications

The reciprocals of the days counted express the observed rates of the insects and mites from egg to adult stage. The range of the observed rates are at [0, 1] interval. The zero value indicates that no development is observed. This situation implies the existence of a truncation point which gives explicitly an upper bound for the upper thermal threshold.

Two datasets are used in this analysis. They concern the study of the two-spotted mite, *Tetranychus urticae* (Barber et al. 2003) and the fourteen-spotted ladybird beetle, *Propylea quatuordecimpunctata* (Papanikolaou et al. 2013). The *Tetranychus urticae* data developmental rates have minimum 0.019 at 15^oC and maximum 0.182 at 32.5^oC. The *Propylea quatuordecimpunctata* dataset consist of 105 beetles and their develop-

Table 3.2 Model selection criteria for the eight models applied to the *Tetranychus urticae* data.

		AIC	DIC	LooCV	WAIC	BIC
Gaussian models	Bieri	-1663.2	-1667.7	-1672.0	-1672.1	-1638.6
	Briere	-1592.2	-1595.8	-1595.8	-1595.9	-1574.7
	Analytis	-1738.3	-1736.9	-1744.0	-1744.0	-1717.3
	Lactin	-1673.9	-1675.6	-1689.1	-1689.1	-1638.8
Inverse Gamma models	Bieri	-1715.6	-1716.0	-1717.2	-1717.2	-1698.1
	Briere	-1749.4	-1749.4	-1746.3	-1746.3	-1735.4
	Analytis	-1889.0	-1887.3	-1889.0	-1889.1	-1867.6
	Lactin	**-1911.4**	**-1911.0**	**-1910.5**	**-1910.6**	**-1893.9**
		$log\,(P_yIS)^{\dagger}$ (se)	$log\,(P_yPP)^{\ddagger}$ (se)		$log\,(P_yBS)^{\S}$ (se)	
Gaussian models	Bieri	796.1 (28.6)	777.2 (19.0)		792.6 (7.1)	
	Briere	759.6 (39.0)	728.2 (12.4)		758.0 (6.9)	
	Analytis	815.9 (40.5)	830.2 (11.0)		821.6 (12.3)	
	Lactin	797.8 (40.0)	821.8 (23.2)		793.8 (56.4)	
Inverse Gamma models	Bieri	837.4 (42.6)	839.7 (9.0)		832.1 (10.0)	
	Briere	846.7 (46.1)	874.0 (12.8)		848.0 (3.5)	
	Analytis	904.2 (42.6)	928.2 (16.3)		903.9 (26.2)	
	Lactin	**920.9 (30.3)**	**956.4 (17.0)**		**922.0 (6.5)**	

† $log\,(P_yIS)$ denotes the logarithm of estimated marginal likelihood via Importance sampling,
‡ $log\,(P_yPP)$ denotes the logarithm of estimated marginal likelihood via Power posterior,
§ $log\,(P_yBS)$ denotes the logarithm of estimated marginal likelihood via Bridge sampling.

mental rate have minimum 0 at 35^oC and maximum 0.111 at 32.5^oC.

3.5.1 Tetranychus urticae example

3.5.1.1 Experimental design

There are 247 mites that have reached adult stage until the study ended. The four ecological models are used both assuming the Gaussian and the Inverse Gamma distributions for the data. The information criteria along with the estimates of the marginal likelihood for each model are provided in Table

3.5.1.2 Estimation of the parameters

3.5.1.3 Model comparison

Both Information criteria and marginal likelihood results clearly suggest that the Inverse Gamma distribution has better fit than the Gaussian distribution at *Tetranychus urticae* dataset across all the ecological models. Furthermore the Lactin model with the Inverse

Gamma distribution stands out against all the other cases. In the Gaussian case, the Analytis model excels.The use of the Inverse Gamma distribution, on the other hand, not only increases the Analytis model's efficiency but also adds flexibility to the Lactin model. The Briere model has the poorest criteria values (Table 3.2).

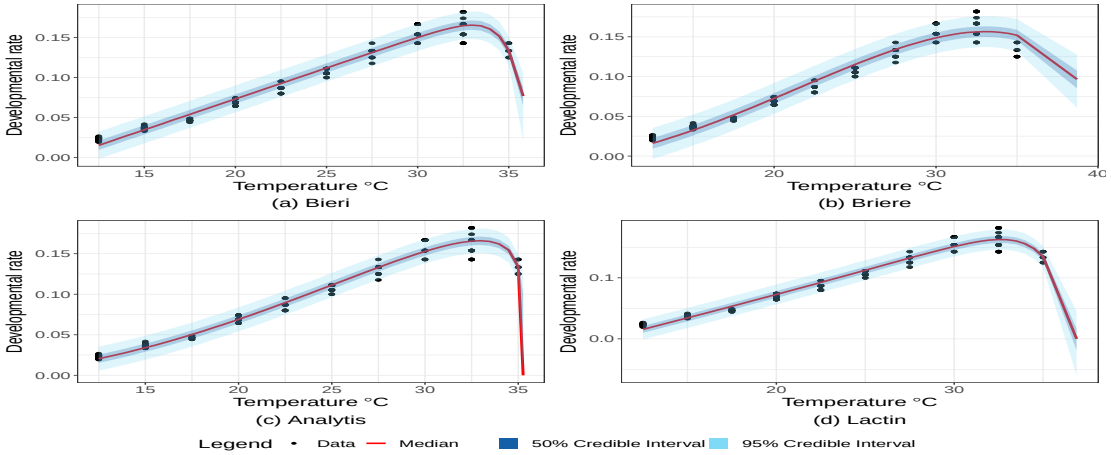

Figure 3.5 Posterior predictive distributions versus *Tetranychus urticae* data using Gaussian distribution.

Although the information criteria have fairly indicated the Briere model as the most suitable for the data, when we concentrate on the upper thermal threshold T_{max} estimate, its performance is poor compared to the other ecological models shown in Fig. 3.6. Nevertheless, even though there is a clear picture concerning information criteria values between the ecological models, there is some variation between marginal likelihood estimates within ecological models in Table 3.2.

Posterior means, 95% credible limits and the effective sample size (neff) of the thermal thresholds and the deviance of the four ecological models are summarised in Tables 3.7 and 3.3 using the Inverse Gamma and the the Gaussian distribution, respectively. In addition, the HMC, ADVI meanfield and ADVI fullrank methods appear alternately in each column for each parameter of interest.

The T_{min} credible limits estimates between the four ecological models do not overlap in Tables 3.7 and 3.3. Bieri model has greater limits and Lactin model gives negative value estimates in the Inverse Gamma case Table 3.7. The credible limits for T_{opt} overlap between Bieri and Analytis whereas the estimates are lower for Lactin model and greater for Briere model. The T_{max} credible limits overlap for Bieri and Lactin in Table 3.3, has higher values for Briere and lower for Analytis model.

3.5.1.4 Computational methods comparison

The credible limits using ADVI meanfield and fullrank methods overlap with the HMC credible limits in most cases in Tables 3.7 and 3.3. Also the ADVI fullrank method

Table 3.3 Posterior summaries for the four models using the Gaussian distribution for the *Tetranychus urticae* data. In each column we report the HMC, the ADVI-Mean field and ADVI-Full rank estimates respectively.

		HMC	ADVI meanfield	ADVI fullrank	neff‡
Bieri	T_{min}	10.5 (10.2, 10.8)	9.8 (9.5, 10.2)	9.8 (9.2, 10.3)	15089
	T_{opt}	33.0 (32.7, 33.4)	158.6 (144.4, 165.4)	183.8 (76.4, 368.5)	8164
	T_{max}	36.3 (35.9, 36.8)	164.9 (161, 169)	190.6 (80, 376.1)	7561
	dev †	-1677.1 (-1683.2, -1666.6)	-1461.0 (-1467.7, -1443.6)	-1461.7 (-1467.7, -1456.3)	10087
Briere	T_{min}	9.3 (8.6, 9.9)	9.3 (9.1, 9.5)	9.3 (8.6, 9.9)	8690
	T_{opt}	33.1 (32.5, 33.7)	33.0 (32.8, 33.2)	33.1 (32.4, 33.7)	8028
	T_{max}	40.0 (39.3, 40.9)	39.9 (39.6, 40.1)	40.0 (39.1, 40.9)	7562
	dev †	-1602.2 (-1606.8, -1593.4)	-1593.2 (-1605.7, -1569.1)	-1598.5 (-1606.3, -1583.6)	9618
Analytis	T_{min}	4.4 (4.0, 5.3)	4.3 (4.2, 4.4)	5.3 (4.8, 6.1)	8757
	T_{opt}	32.9 (32.7, 33.3)	33.0 (32.9, 33.1)	33.2 (32.6, 33.7)	41
	T_{max}	35.3 (35.1, 35.5)	35.2 (35.2, 35.3)	35.2 (35.1, 35.4)	8246
	dev †	-1750.3 (-1758.0, -1738.3)	-1744.7 (-1755.6, -1724)	-1709.1 (-1741.9, -1606.9)	10954
Lactin	T_{min}	10.4 (10.1, 10.7)	-3.9 (-5.8, -1.9)	-0.7 (-4.0, 3.0)	11526
	T_{opt}	32.6 (32.4, 32.8)	32.2 (32.0, 32.3)	32.0 (31.8, 32.2)	9884
	T_{max}	36.9 (36.6, 37.3)	38.3 (38.1, 38.5)	38.2 (37.9, 38.5)	7797
	dev †	-1693.9 (-1703, -1679.6)	-1764.0 (-1780.3, -1738.1)	-1776.2 (-1784.9, -1762)	8227

† deviance of the model given the data,
‡ effective sample size.

Table 3.4 BMA weights for the *Tetranychus urticae* data.

		Bieri	Briere	Analytis	Lactin
Gaussian distribution	aic_w	4.9E-17	1.9E-32	$0.\overline{9}$	1.0E-14
	dic_w	9.4E-16	2.3E-31	$0.\overline{9}$	4.9E-14
	loocv_w	2.3E-16	6.6E-33	$0.\overline{9}$	1.2E-12
	waic_w	2.4E-16	6.9E-33	$0.\overline{9}$	1.2E-12
	bic_w	8.1E-18	1.1E-31	$0.\overline{9}$	9.0E-18
	elbo_f	3.1E-53	2.9E-27	0.001	0.999
	elbo_r	7.3E-45	1.5E-18	$0.\overline{9}$	3.5E-39
Inverse Gamma distribution	aic_w	**3.0E-43**	6.6E-36	1.4E-5	$0.\overline{9}$
	dic_w	4.5E-43	8.1E-36	7.1E-6	$0.\overline{9}$
	loocv_w	1.1E-42	2.2E-36	2.1E-5	$0.\overline{9}$
	waic_w	1.0E-42	2.1E-36	2.1E-5	$0.\overline{9}$
	bic_w	3.0E-43	3.8E-35	1.9E-6	$0.\overline{9}$
	elbo_f	1.6E-20	8.9E-19	$0.\overline{9}$	1.0E-36
	elbo_r	1.0E-200	1.0E-171	$0.\overline{9}$	2.7E-58

seems to be closer to the HMC estimates as in Briere and Lactin models in Table 3.7. However in general it has worst fit than the corresponding fitted model using HMC and also gives wider 95% Cr.I..

3.5.1.5 BMA performance

Bayesian model averaging provides alternative estimates for the parameters of interest, combining the predictive efficiency of all four ecological models. The derived weights are shown in Table 3.4 whereas the BMA estimates and their 95% Cr.I. are given on Table 3.5 and are divided into the data case of the Gaussian distribution and the data case of the Inverse Gamma distribution. The predictive bias of the Analytis and Lactin models appear to affect the model averaging estimates of T_{min}, T_{max} and the deviance in the Gaussian and the Inverse Gamma case respectively.

Also different weights based on (3.20) give almost identical 95% credible limits. However using ELBO based BMA weights do not give robust estimates of the parameters of interest which is not unexpected since the ELBO is a lower bound estimate of the marginal likelihood. Furthermore, the time elapsed until the completion of the algorithm for ADVI methods is up to 52 seconds, while for the HMC method is at least 476 seconds for Tetranychus dataset as shown in Table 3.6.

Table 3.5 The mean and 95% Cr.I. limits of the BMA estimates of parameters of interest calculated with the use of information criteria score weights assuming the Gauussian and Inverse Gamma distributions respectively for *Tetranychus urticae* data.

		T_{min}	T_{opt}	T_{max}	dev†
		Mean 95% Cr.I.	Mean 95% Cr.I.	Mean 95% Cr.I.	Mean 95% Cr.I.
Gaussian Model	aic_w	4.4 (4.0, 5.3)	32.9 (32.7, 33.3)	35.3 (35.1, 35.5)	-1750.3 (-1758.0, -1738.3)
	dic_w	4.4 (4.0, 5.3)	32.9 (32.7, 33.3)	35.3 (35.1, 35.5)	-1750.3 (-1758.0, -1738.3)
	loocv_w	4.4 (4.0, 5.3)	32.9 (32.7, 33.3)	35.3 (35.1, 35.5)	-1750.3 (-1758.0, -1738.3)
	waic_w	4.4 (4.0, 5.3)	32.9 (32.7, 33.3)	35.3 (35.1, 35.5)	-1750.3 (-1758.0, -1738.3)
	bic_w	4.4 (4.0, 5.3)	32.9 (32.7, 33.3)	35.3 (35.1, 35.5)	-1750.3 (-1758.0, -1738.3)
	elbo_mf	4.4 (4.0, 5.3)	32.6 (32.4, 32.8)	36.9 (36.5, 37.3)	-1693.9 (-1703.1, -1679.7)
	elbo_fr	4.4 (4.0, 5.3)	32.9 (32.7, 33.3)	35.3 (35.1, 35.5)	-1750.3 (-1758.0, -1738.3)
Inverse Gamma Model	aic_w	4.2 (4.0, 4.6)	32.0 (31.8, 32.2)	38.4 (38.1, 38.9)	-1916.3 (-1920.6, -1908.3)
	dic_w	4.2 (4.0, 4.6)	32.0 (31.8, 32.2)	38.4 (38.1, 38.9)	-1916.3 (-1920.6, -1908.3)
	loocv_w	4.2 (4.0, 4.6)	32.0 (31.8, 32.2)	38.4 (38.1, 38.9)	-1916.3 (-1920.6, -1908.3)
	waic_w	4.2 (4.0, 4.6)	32.0 (31.8, 32.2)	38.4 (38.1, 38.9)	-1916.3 (-1920.6, -1908.3)
	bic_w	4.2 (4.0, 4.6)	32.0 (31.8, 32.2)	38.4 (38.1, 38.9)	-1916.3 (-1920.6, -1908.3)
	elbo_mf	4.2 (4.0, 4.6)	33.6 (33.3, 33.9)	35.0 (35.0, 35.1)	-1894.0 (-1899.2, 1885.2)
	elbo_fr	4.2 (4.0, 4.6)	33.6 (33.3, 33.9)	35.0 (35.0, 35.1)	-1894.0 (-1899.2, 1885.2)

† deviance of the model given the data.

Table 3.6 Algorithmic working time in seconds, for two datasets.

Tetranychus urticae

Gaussian				
	Bieri	Briere	Analytis	Lactin
HMC	2011	476	9852	5672
ADVI-meanfield	8	1	52	1
ADVI-fullrank	45	6	38	44
Inverse Gamma				
HMC	1375	761	3231	4917
ADVI-meanfield	2	4	15	27
ADVI-fullrank	2	34	14	44

Propylea quatuordecimpunctata

Gaussian				
	Bieri	Briere	Analytis	Lactin
HMC	4335	20218	4913	1426
ADVI-meanfield	13	7	3	16
ADVI-fullrank	48	11	12	16
Zero Inflated Inverse Gamma				
HMC	1078	321	15291	3073
ADVI-meanfield	41	5	9	10
ADVI-fullrank	52	5	39	18

The average difference between the working times of HMC and each ADVI method is around 3500 seconds (58 minutes) whereas between the meanfiled and fullrank is around 18 seconds. These findings illuminate the VBI fullrank method's time-efficacy.

In Fig. 3.6 the posterior predictive distributions versus *Tetranychus urticae* data using Inverse Gamma distribution are shown for the ecological models.

The adaptivity in data across predictor values is clear in all the ecological models, where the 95% credible limits are adjusted to data variance in each temperature level. On the other hand, in the Gaussian case (Fig. 3.5) the variance of the posterior remains constant across predictor values.

Table 3.7 Posterior summaries for the four models using the Inverse Gamma distribution for the *Tetranychus urticae* data. In each column we report the HMC, the ADVI-meanfield and ADVI-fullrank estimates respectively.

		HMC	ADVI meanfield	ADVI fullrank	neff‡
Bieri	T_{min}	9.4 (9.3, 9.6)	9.3 (9.1, 9.4)	9.3 (9.0, 9.5)	19741
	T_{opt}	33.7 (33.2, 34.1)	145.5 (134.8, 149.8)	203.2 (114.5, 326.6)	8888
	T_{max}	35.8 (35.4, 36.5)	150.8 (149.2, 152.5)	209.0 (114.5, 326.6)	7738
	dev †	-1720.9 (-1724.9, -1713)	-1652.9 (-1658.6, -1637.6)	-1652.4 (-1658.5, -16)	10228
Briere	T_{min}	6.6 (6, 7)	6.6 (6.4, 6.7)	6.5 (6.1, 7)	7933
	T_{opt}	36.7 (35.2, 38.5)	36.6(36.1, 37)	36.6 (35, 38.3)	6978
	T_{max}	45.0 (43.1, 47.3)	44.8 (44.2, 45.4)	44.9 (42.8, 47)	6935
	dev †	-1753.3(-1757, -1745.9)	-1752.9 (-1757.1, -1741.4)	-1751.9 (-1756.7, -1740.4)	8638
Analytis	T_{min}	4.2 (4, 4.6)	7.6 (7.5, 7.8)	6.0 (4.4, 9.4)	9977
	T_{opt}	33.6 (33.3, 34)	99.9 (96.6, 103)	79.9 (16.8, 239.3)	7876
	T_{max}	35.0 (35, 35.1)	99.9(96.6, 103.1)	80.9 (17.5, 240.6)	9038
	dev †	-1894.0 (-1899.2, -1885)	-1670.9 (-1691.8, -1641.4)	-1495.1 (-1687.4, -149.5)	12370
Lactin	T_{min}	-18.7 (-18.7, -18.7)	7.8 (7.6, 8.1)	-18.7 (-18.8, -18.7)	18393
	T_{opt}	32.0 (31.8, 32.2)	32.1 (31.8, 32.3)	33.9 (33.8, 34.1)	12228
	T_{max}	38.4 (38.1, 38.9)	42.6 (42.4, 42.8)	38.6 (38.2, 39)	8284
	dev †	-1916.3 (-1920.6, -1908.3)	-1705.8 (-1747.8, -1650.5)	-1890.0 (-1918.5, -1791.8)	9938

† deviance of the model given the data.

‡ effective sample size.

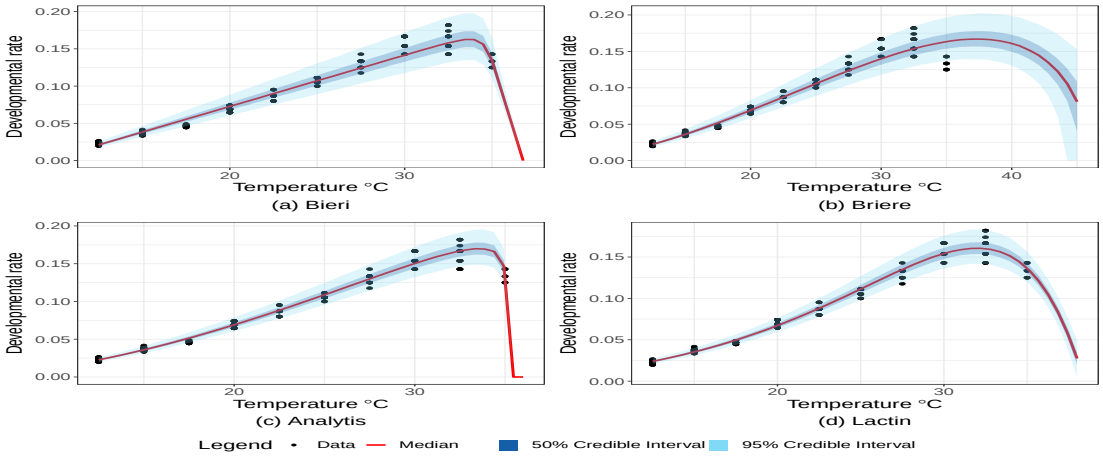

Figure 3.6 Posterior predictive distributions versus *Tetranychus urticae* data using Inverse Gamma distribution.

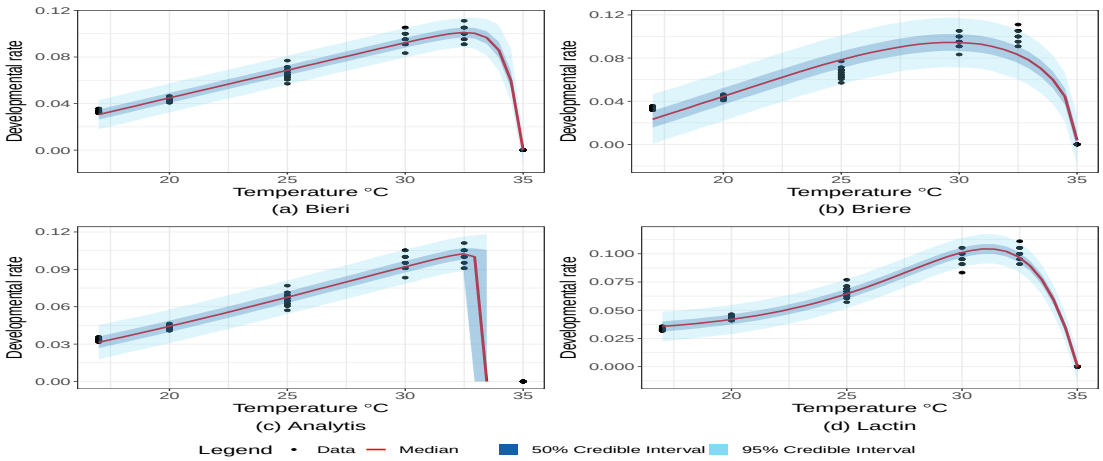

Figure 3.7 Posterior predictive distributions versus *Propylea Coccinellidae* data using Gaussian distribution.

3.5.2 *Propylea quatuordecimpunctata* example

3.5.2.1 Experimental design

There are 17 out of 105 insects that have not altered their egg status until the study ended. These cases are observed at $35°C$ and are indicated by zeros in the response variable y. The Inverse Gamma distribution is not defined for zero response values. On the other hand, the Bieri and Lactin models can also generate negative values. Hence, in order to account for the presence of zeros, we use a Zero Inflated Inverse Gamma model and choose parameter initial values so as the scale of the Inverse Gamma to remain positive. Both the Gaussian and the Zero Inflated Inverse Gamma distributions are used for the data along with the four ecological models. The various model selection tools are summarised in Table 3.8.

Table 3.8 Model selection criteria for the eight models applied to the *Propylea Coccinellidae* data.

		AIC	DIC	LooCV	WAIC	BIC
Gaussian models	Bieri	**-804.7**	**-783.6**	**-816.1**	**-816.1**	**-786.1**
	Briere	-650.5	-649.5	-658.3	-658.3	-637.3
	Analytis	-662.1	-637.1	-671.4	-671.4	-646.1
	Lactin	-789.8	-776.8	-806.7	-806.7	-763.3
Inverse Gamma model	Bieri	-716.1	-719.3	-719.3	-719.3	-702.8
	Briere	-560.8	-561.4	-561.4	-561.4	-550.2
	Analytis	-734.7	-732.4	-732.4	-732.4	-718.8
	Lactin	-698.9	-699.1	-699.1	-699.1	-685.6
		$log\,(P_yIS)^{\dagger}$ (se)	$log\,(P_yPP)^{\ddagger}$ (se)		$log\,(P_yBS)^{\S}$ (se)	
Gaussian models	Bieri	**419.9 (19.6)**	**433.4 (18.0)**		**462.4 (9.2)**	
	Briere	299.2 (24.6)	235.4 (6.7)		231.4 (1.2)	
	Analytis	283.7 (24.0)	312.2 (10.0)		274.2 (33.7)	
	Lactin	390.1 (33.5)	328.1 (23.0)		334.7 (1.7)	
Inverse Gamma models	Bieri	334.5 (25.9)	342.9 (31.5)		333.2 (5.0)	
	Briere	268.2 (23.2)	265.7 (4.6)		251.7 (1.9)	
	Analytis	340.8 (26.7)	371.3 (14.5)		335 (31.2)	
	Lactin	329.2 (19.5)	329.3 (11.8)		313.7 (18.8)	

† $log\,(P_yIS)$ denotes the logarithm of estimated marginal likelihood via Importance sampling,
‡ $log\,(P_yPP)$ denotes the logarithm of estimated marginal likelihood via Power posterior,
§ $log\,(P_yBS)$§denotes the logarithm of estimated marginal likelihood via Bridge sampling.

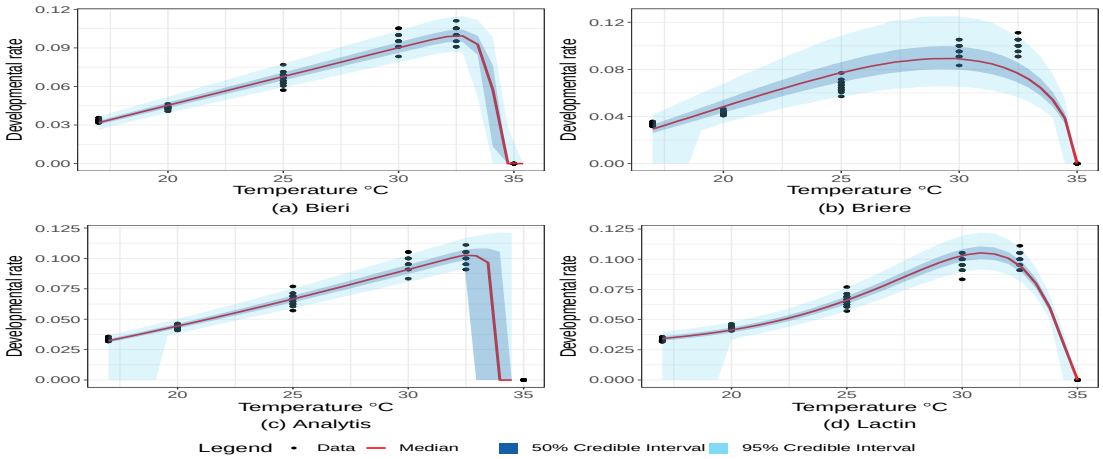

Figure 3.8 Posterior predictive distributions versus *Propylea Coccinellidae* data using Zero Inflated Inverse Gamma distribution.

Both Information criteria and marginal likelihood results clearly suggest that the Zero Inflated Inverse Gamma distribution has better fit than the Gaussian distribution at Propylea quatordicempuncata only at the Analytis model. Furthermore Bieri and Analytis models stand out in the Gaussian and the Inverse Gamma case respectively, whilst the Briere model has the poorest criteria values (Table 3.8). Nevertheless, even though there is a clear picture of the goodness of fit between the ecological models, there is some variation between marginal likelihood estimates within ecological models in Table 3.8.

3.5.2.2 Model comparison

In the Gaussian model, the Bieri and Lactin models stand out according to the information criteria and marginal likelihood estimates, the Analytis model follows, whilst the Briere model has the lower criteria values. On the contrary, in the Zero Inflated Inverse Gaussian model, the Analytis model is a better choice according to the marginal likelihood and the information criteria values, while Bieri, Lactin and Briere models follow respectively. Lactin and Bieri can be interpreted by their ability to track an exponential data-value decrease. Additionally, the Bieri model manages to capture the linear increase in the *Propylea quatuordecimpunctata* dataset. The Briere model, on the other hand, lacks performance due to its unique structure, which requires that the decline of the response variable as the temperature rises be of the form $\sqrt{T_{max} - T}$. The Analytis model has similar multiplicative structure to the Briere model, but it is more complex model since the exponents of its model are unknown variables that make it adaptive and liable to data change especially in the Inverse Gamma case.

Table 3.9 Posterior summaries for the four models using the Gaussian distribution for the *Propylea Coccinellidae* data. In each column we report the HMC, the ADVI-Mean field and ADVI-Full rank estimates respectively.

		HMC	ADVI meanfield	ADVI fullrank	neff‡
Bieri	T_{min}	10.6 (9.7, 11.4)	10.7 (10.6, 10.8)	10.7 (10, 11.4)	7906
	T_{opt}	32.6 (32, 33.5)	32.2 (32, 32.5)	32.2 (31.8, 32.5)	7620
	T_{max}	35.0 (34.98, 35.02)	33.8 (33.6, 33.9)	33.7 (33.5, 33.9)	7194
	dev †	-818.7(-833.1, -800.6)	-672.7 (-685.1, -657.4)	-672.5 (-686.2, -657.1)	11673
Briere	T_{min}	13.1 (12.4, 13.8)	13.3 (12.9, 13.8)	8.2 (1.8, 15.6)	4
	T_{min}	29.7 (29.6, 29.8)	29.7 (29.5, 29.8)	27.3 (11.7, 30.1)	4
	T_{min}	35.0 (35.0, 35.01)	35.0 (34.8, 35)	32.9 (14.3, 35)	113
	dev †	-660.5 (-667.0, -649.7)	-535.8 (-543.2, -520.5)	-429.7 (-535.0, -402.1)	14478
Analytis	T_{min}	6.3 (4.1, 10.6)	21.5 (7.7, 55.2)	6.9 (5.4, 9.3)	3742
	T_{min}	33.2 (32.1, 34.8)	39.2 (18.1, 89.1)	33.2 (31.7, 37.7)	4435
	T_{min}	33.6 (32.5, 34.9)	42.5 (32.7, 94.5)	36.1 (4.6, 133.2)	9370
	dev †	-674.1 (-690.2, -666.3)	-160.0 (-261.3, -98.2)	-491.9 (-595.9, -293.6)	5164
Lactin	T_{min}	-154.4 (-234.7, -119.3)	-146.2 (-154.3, -138.1)	-159.2 (-243.1, -119.0)	7560
	T_{min}	31.2 (30.9, 31.3)	31.1 (31.1, 31.2)	31.1 (30.9, 31.3)	8024
	T_{min}	35.0 (35, 35.04)	35.0 (35.0, 35.1)	35.0 (35.0, 35.1)	21622
	dev †	-809.8 (-823.7, -792.2)	-804.7 (-820.5, -784.8)	-778.5 (-807.5, -702.6)	12147

† deviance of the model given the data,
‡ effective sample size.

3.5.2.3 Estimation of the parameters

Nevertheless, the Analytis model needs some tuning when defining its thermal parameters space. This is necessary to avoid allowing T_{min} values close to zero, which would result in parameter underestimation and poor fit in the Analytis model. Especially, in the case of the Propylea dataset, we a-prior assumed that T_{min} is greater than $4°C$. Posterior means, 95% credible limits and the effective sample size of the thermal thresholds and the deviance of the four ecological models are summarised in Tables 3.10 and 3.9 using the Zero inflated Inverse Gamma and the Gaussian distribution respectively. Specifically, the HMC, ADVI meanfield and ADVI fullrank methods appear alternatively in each column of Tables (3.10 and 3.9) for each parameter of interest. Briere model has greater 95% credible limits for T_{min} while Lactin model gives negative value estimates. The results for T_{opt} overlap for Bieri and Analytis in the Zero Inflated Inverse Gamma case whereas the estimates are lower for Lactin model. The T_{max} credible limits overlap between Bieri and the other ecological models both in the Gaussian and the Zero Inflated Inverse Gamma case. The zero-rate values can be naturally modelled in the case of the Gaussian distribution. Though only the Bieri and Analytis models give wider 95% Cr.I.. For the rest of the models, the credible limits difference is lower than three degrees $3°C$. In addition, in Zero Inflated Inverse Gamma case, the Analytis model has higher information criteria values, while the Bieri and Lactin models follow as shown in Table 3.8, respectively. All four ecological models incorporate terms into their structure in the form of products which directly affect the scale of the Inverse Gamma distribution. Thus, not only the mean but also the variance of the statistical model change along and adapt the flunctuations for each temperature level as shown in Fig. 3.8. This adaptivity is evident across all the four ecological models in use. Even though no specific mechanism is used to account for excess zeros and the variance of the posterior predictive remains constant across predictor values in the Gaussian case, the fit to *Propylea Coccinellidae* data is good, as shown in Figure 3.7.

The T_{min} outcomes for 95% credible limits do not overlap, while Briere model has larger values. The Lactin model is not suitable to provide reasonable estimates as it allows for negative values (Table 3.10). The results for T_{opt} are sorted in ascending order from Bieri, Lactin, Analytis, and Briere. We observe that the credible limits for the T_{max} overlap between models and do not exceed the maximum observed temperature 35^oC. Concerning the Zero Inflated Inverse Gamma case, the posterior predictive probabilities of non-zero entries, which are the probabilities of anthropods' development, are shown in Figure 3.9. We observe that the probabilities tend to unity near T_{opt} estimates whereas become negligible out of the thermal threshold limits except for the Lactin model case which does not give robust estimates for T_{min}.

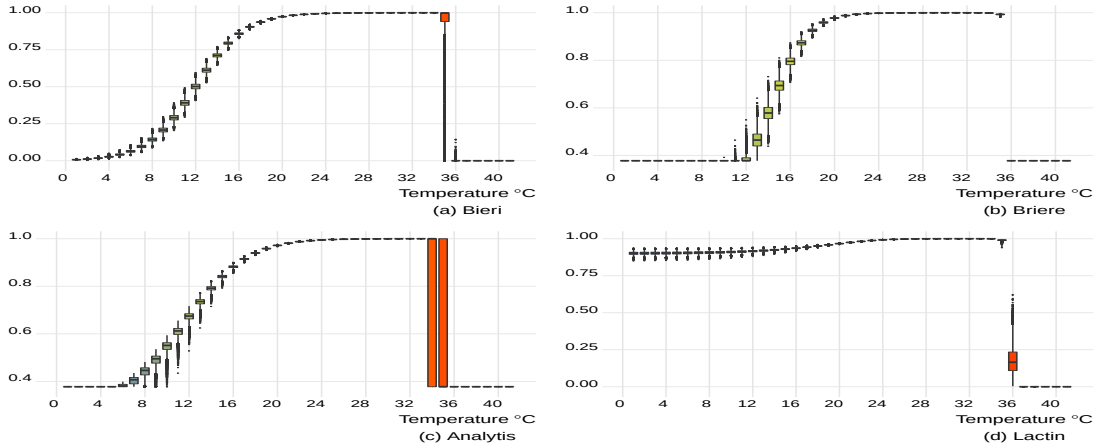

Figure 3.9 Boxplots of the posterior probability of non-zero development in *Propylea Coccinellidae* data using the Zero Inflated Inverse Gamma model.

3.5.2.4 Computational methods comparison

In addition, the 95% credible limits estimates using ADVI meanfield and fullrank methods overlap with the HMC estimates with exceptions at the Lactin model in which the fullarank method seems to be closer to the HMC estimates (Tables 3.9 and 3.10). In many cases, the ADVI fullrank method agrees with HMC. It seems to give more robust estimates for T_{max} mean in Gaussian case and Bieri model. However it has worst fit than HMC and also gives wider 95% Cr.I..

Furthermore, the time elapsed until the completion of the algorithm for ADVI methods is up to 48 seconds, while for the HMC method it is at least 321 seconds for Propylea dataset as shown in Table 3.6. The average difference between the working times of HMS and each ADVI method is around 6300 seconds (105 minutes) whereas between the meanfield and fullrank is around 12 seconds. These findings illuminate both the VBI fullrank method's computational time-efficacy.

3.5.2.5 BMA performance

Bayesian model averaging results are shown in Tables 3.11, 3.12. This time the predictive bias that affect the model averaging estimates are from the Bieri and Analytis models for the Gaussian and the Zero Inflated Inverse Gamma distribution respectively.

3.6 Chapter summary

There are several issues that we deal with in terms of arthropod developmental rates, and there are some concerns for future research. To begin with, comparing non-linear non-nested models with varying numbers of parameters and truncated mean structures, as well as excessive-zeros in data, is not a trivial task.

Table 3.10 Posterior summaries for the four models using the Zero Inflated Inverse Gamma distribution for the *Propylea Coccinellidae* data. In each column we report the HMC, the ADVI-Mean field and ADVI-Full rank estimates respectively.

		HMC	ADVI meanfield	ADVI fullrank	neff‡
Bieri	T_{min}	9.9 (9.4, 10.3)	9.4 (9.3, 9.6)	9.2 (8.7, 9.7)	9138
	T_{opt}	32.7 (32.2, 33.8)	73.0 (32.6, 146.6	109.4 (28.7, 320.1)	16552
	T_{max}	34.4 (33.7, 34.9)	18.1 (9.3, 52.6)	16.0 (8.8, 52.5)	12972
	dev †	-722.3 (-725.7, -714.8)	-318.5 (-360.4, -318.8)	-323.8 (-572.1, -309.3)	9962
Briere	T_{min}	11.1 (10.3, 11.9)	11.2 (10.8, 11.5)	11.2 (10.4, 12.0)	14192
	T_{opt}	29.3 (29.2, 29.5)	29.3 (29.2, 29.4)	29.3 (29.2, 29.5)	17300
	T_{max}	35.0 (34.8, 35.0)	34.9 (34.8, 35.0)	34.8 (34.8, 35.0)	23972
	dev †	-563.8 (-568.0, -555.9)	-563.1 (-567.7, -552.0)	-561.9 (-567.4, -549.7)	14077
Analytis	T_{min}	5.0 (4.0, 7.0)	5.1 (4.1, 8.7)	4.4 (4.2, 4.9)	352
	T_{opt}	33.5 (32.3, 34.9)	33.0 (29.71, 37.0)	32.2 (30.5, 33.9)	1240
	T_{max}	33.6 (32.5, 34.9)	33.6 (30.3, 37.6)	32.5 (30.7, 34.4)	1409
	dev †	-736.2 (-742.1, -729.1)	-411.6 (-482.3, -216.9)	-684.5 (-742.5, -618.8)	1117
Lactin	T_{min}	-133.51 (-157.5, -116.8)	11.6 (1.2, 59.2)	-12.5 (-409.8, 8.5)	11511
	T_{opt}	30.9 (30.8, 31.1)	39.7 (1.1, 211.2)	25.9 (5.5, 77.7)	9847
	T_{max}	34.9 (34.7, 35.0)	25.3 (1.1, 76.3)	-45.1 (-935.4, 52.5)	16361
	dev †	-703.6 (-708.1, -695.0)	247.8 (0.0, 294.8)	247.8 (0.0, 294.8)	9179

† deviance of the model given the data,
‡ effective sample size.

Table 3.11 BMA weights for the *Propylea quatuordecimpunctata* data.

		Bieri	Briere	Analytis	Lactin
Gaussian distribution	aic_w	$0.\overline{9}$	3.3E-34	1.1E-31	5.9E-04
	dic_w	0.968	7.3E-30	1.5E-32	0.032
	loocv_w	0.991	5.5E-35	3.8E-32	0.009
	waic_w	0.991	5.5E-35	3.8E-32	0.009
	bic_w	$0.\overline{9}$	4.8E-33	4.0E-31	1.1E-05
	elbo_mf	$0.\overline{9}$	5E-203	9.0E-269	5.0E-159
	elbo_fr	$0.\overline{9}$	3E-218	1.0E-210	3.0E-162
Zero Inflated Inv. Gamma distribution	aic_w	9.3E-5	1.7E-38	$0.\overline{9}$	1.7E-8
	dic_w	0.001	3.9E-38	$0.\overline{9}$	2.1E-8
	loocv_w	0.001	7.3E-38	$0.\overline{9}$	5.8E-8
	waic_w	0.001	7.2E-38	$0.\overline{9}$	5.8E-8
	bic_w	3.5E-4	2.4E-37	$0.\overline{9}$	6.4E-8
	elbo_mf	5.1E-53	$0.\overline{9}$	3.3E-35	3.0E-114
	elbo_fr	3.4E-72	1.1E-23	$0.\overline{9}$	2.0E-140

3.6.1 Computational methods overview

We use the Bayesian paradigm, as well as some contemporary computational approaches such as the HMC, ADVI-meanfield, and ADVI-fullrank, to address not only irregular and truncated mean structures, but also the uncertainty of zero generation in the data. Although ADVI techniques are gaining popularity in the scientific community due to their fast and computationally inexpensive approximations to the posterior distributions, they do not provide robust estimates of all the parameters of the models we study. The HMC method, on the other hand, gives robust estimates even under these specific model and data structure conditions.

3.6.2 Distribution of the data

Furthermore, for the data generation scheme, we suggest the Gaussian and the Inverse Gamma distributions. The Gaussian option provides sensible estimates that can be used even though the data has a lot of zeros. Inverse Gamma, on the other hand, not only naturally models developmental rates, which are characterized as the reciprocal of positive real values, but also provides variance adaptivity across temperature fluctuations. Also for each ecological model we define the Zero Inflated Inverse Gamma density so as to model data with an excessive number of zeros. When comparing models involving Gaussian and Inverse Gamma or Zero Inflated gamma distributions, we find that the second performs better in non-zero data cases, while the first performs better in all but the Analytis model.

Table 3.12 The mean and 95% Cr.I. limits of the BMA estimates of parameters of interest calculated with the use of information criteria score weights assuming the Gaussian and Inverse Gamma distributions respectively for the *Propylea quatuordecimpunctata* data.

		T_{min}	T_{opt}	T_{max}	dev†
		Mean 95% Cr.I.	Mean 95% Cr.I.	Mean 95% Cr.I.	Mean 95% Cr.I.
Gaussian distribution	aic_w	10.6 (9.7, 11.4)	32.6 (32.0, 33.5)	35.0 (34.98, 35.0)	-818.8 (-833.1, -800.8)
	dic_w	10.6 (9.7, 11.4)	32.5 (21.9, 33.4)	35.0 (34.98, 35.0)	-818.5 (-832.4, -801.1)
	loocv_w	10.6 (9.7, 11.4)	32.6 (32.0, 33.5)	35.0 (34.98, 35.0)	-818.7 (-832.9, -800.9)
	waic_w	10.6 (9.7, 11.4)	32.6 (32.0, 33.5)	35.0 (34.98, 35.0)	-818.7 (-832.9, -800.9)
	bic_w	10.6 (9.7, 11.4)	32.6 (32.0, 33.5)	35.0 (34.98, 35.0)	-818.8 (-833.1, -800.8)
	elbo_mf	10.6 (9.7, 11.4)	32.6 (32.0, 33.5)	35.0 (34.98, 35.0)	-818.8 (-833.1, -800.8)
	elbo_fr	10.6 (9.7, 11.4)	32.6 (32.0, 33.5)	35.0 (34.98, 35.0)	-818.8 (-833.1, -800.8)
Zero Inflated Inv. Gamma distribution	aic_w	5.0 (4.0, 7.0)	33.5 (32.3, 34.9)	33.6 (32.5, 34.9)	-736.2 (-742.1, -729.1)
	dic_w	5.0 (4.0, 7.0)	33.5 (32.3, 34.9)	33.6 (32.5, 34.9)	-736.2 (-742.1, -729.1)
	loocv_w	5.0 (4.0, 7.0)	33.5 (32.3, 34.9)	33.6 (32.5, 34.9)	-736.2 (-742.1, -729.1)
	waic_w	5.0 (4.0, 7.0)	33.5 (32.3, 34.9)	33.6 (32.5, 34.9)	-736.2 (-742.1, -729.1)
	bic_w	5.0 (4.0, 7.0)	33.5 (32.3, 34.9)	33.6 (32.5, 34.9)	-736.2 (-742.1, -729.1)
	elbo_mf	11.1 (10.3, 11.9)	29.3 (29.2, 29.5)	35.0 (34.8, 35.0)	-563.8 (-568.0, -555.9)
	elbo_fr	5.0 (4.0, 7.0)	33.5 (32.3, 24.9)	33.6 (32.5, 34.9)	-736.2 (-742.1, -729.1)

† deviance of the model given the data.

3.6.3 Model comparison

In addition, we address the model comparison challenge by employing the Information criteria not only to assess model goodness of fit to data while accounting for model complexity, but also to assess ability of models to make predictions on new data using the leave one out cross validation technique. Additionally, we use marginal likelihood approximations of the various models to determine which one is best supported by the data. Finally, we plot the posterior predictive distributions alongside the observed data points to visualize the prediction ability of the suggested models.

3.6.4 BMA performance

The predictive bias of the weighted models, as well as the uncertainty about the weights, affect BMA results using Information criteria as weights according to (3.20) weights as outlined in section 3.4.2.3 in the appendix. As a result, the BMA approach does not provide estimates that differ from the best-performing models.

3.6.5 Future research

Among things for future research is to select consistently the most robust candidate between models given sufficiently many data samples, in a sensitivity analysis perspective. Moreover, the ADVI methods, can be extended so as to capture more sophisticated mean structures, like the ones we present in the current work. In addition, probability density that generates zeros in the Zero Inflated Inverse Gamma distribution can be modeled in more complex ways, such as using hyperparameters and hierarchical effects across temperature levels. Finally, R-packages that include the suggested models and perform the analysis presented in this Chapter are to be created.

Chapter 4

DRIN: R package for developmental rate inference

4.1 Introduction

Following Chapter 3, mathematical modelling is an indispensable tool for comprehending the ensuing developmental process in all but the simplest cases (Kontodimas et al. 2004). Fitting ecologicalal models for developmental rates, on the other hand, is not straightforward, often because the mathematical forms are not linear (Papanikolaou et al. 2019) and the actual biochemical reactions of insects or environmental factors responsible for their growth may remain unobservable. As a consequence, we use the Bayesian paradigm in conjunction with specific data generation models, as described in Chapter 3. The latter models include not only various data distributions such as the widely used Gaussian, Inverse Gamma, and zero inflated Inverse Gamma, which can also generate excess zeros, but also four popular ecological non-linear functions that describe developmental rate dynamics. Nonetheless, there are significant challenges to overcome, such as the function structure dependence on thermal parameters, parameter ranges, and initial value selection when applying these ecological functions. The various models studied in Chapter 3, as well as various choices for initial values, prior distributions and predictions,are integrated into a generic package in the R platform. The introduced package is called Developmental Ratio Inference (DRIN), and it enables the simple and user-friendly estimation of thermal threshold parameters, which determine the development prospects of various species in ecology. The calculation is based on HMC using STAN functionality. The package is available in GitHub at (Kondakis et al. 2021b) and would be submitted for publication in the Autumn of 2021.

The following sections describe the structure of the 'DRIN' package, including code function notation and arguments used, output format, instructions for embedding data and selecting prior distribution and initial values, and finally the presentation of some examples and chapter discussion.

4.2 Package (DRIN) structure

The package's generic function is denoted *drin_hmc* and it produces the desired outcome. To invoke the *drin_hmc* function, the user must supply several arguments listed below.

- The current data (C_data) argument: is a list containing the data used in the experiment.
- The sampling parameters (sparam) argument: is a vector containing several sampling parameter options.
- The likelihood (lik) argument: specifies the likelihood distribution to be used between the models studied in Chapter 3.
- The ecological model type (mtype) argument: specifies which of the four ecological models studied in Chapter 3 should be used.
- The prior distributions (prior) argument: specifies a set of prior distributions for the parameters of the previously defined ecological model by (mtype).
- The initial values (init_list) argument: specifies a list of options for the initial values of the imposed ecological model defined by (mtype).

4.2.1 Input of data and likelihood function

The (C_data) is a list with multiple elements. The user must supply the observation vectors of the response variable y and the predictor x. If no value is specified, the length of the vector x is used to calculate the number of observations N. Additionally, in elements NP and $xpred$, the user can optionally insert a number or a realization vector of x as predictions to be made by the model imposed. In the absence of the former, the NP is initialized by the number of $xpred$ values, whereas in the absence of the latter, the NP is initialized by a sequence of NP values within the range of x observations. In the event that both are absent the NP is set to the double of the range of predictor observations x.

The (lik) argument accepts the likelihood choice of "gauss" (which is the default choice) and "igamma," which specify the Gaussian and Inverse Gamma distributions, respectively.

4.2.2 Input of parameters and other specifications

The (sparam) is an optional element that specifies the number of chains, iterations, burn-in size, and thinning size to pass information to the R interface of STAN (rstan) function. The sparam element's default values are 4, 11000, 10000, and 1 respectively.

For the Bieri, Briere, Analytis, and Lactin ecological models described in section 3.2, the (mtype) argument accepts the options "bieri" (which is the default), "briere", "analytis", and "lactin" respectively.

4.2.3 Choice of priors

The (prior) element optionally provides the prior distributions of the parameters involved in the ecological model declared by the (mtype) argument to the *drin_hmc* function. The prior distribution declaration can be expressed as a list of parameter names and their corresponding prior distributions, as defined in the STAN manual (Stan Development Team 2021a).For example, in the Lactin ecological model option, the prior distribution of the $del = \frac{1}{\Delta}$ parameter declaration could be changed from the default value to $prior = list(del = "exponential(100)"$. The default options for prior distributions are shown in Table 3.1.

4.2.4 Plug in initial values and other arguments

The (init_list) element optionally provides the *drin_hmc* function with the initial values of the ecological model parameters defined by the (mtype) argument for the shake of the HMC algorithm. As described in Section 3.2, It should be defined as a list or a function of a list containing all model parameters. The initial values declaration can take the form of a list containing the exact parameter names and initial values, or it can take the form of a function of a list containing a random sample from a specific distribution for each parameter. For the Bieri ecological model and the Inverse Gamma likelihood option, for example, the initial value declarations could be either a fixed numeric list

```
init_list=list(a=0.1,tmin=1,b=1.1,tmax=30,shape=500)
```

or a list of functions of randomly generated numbers based on specific distributions for each parameter, taking into account any domain constraints. As an example of the Bieri model, consider the following:

```
inittmin=rgamma(1,shape=10,rate=1)
init_list=function(num_id=1)
{list(a=(1e-1)*rbeta(1,1,1),+
tmin=inittmin*rbeta(1,1,1),+
b=(1e-5)*rbeta(1,1,1),+
tmax=inittmin+rgamma(1,shape=10,rate=1),+
```

```
shape=runif(1,300,500))}
```

In the absence of a (init_list) element, the default initial values used in the application of the *Tetranichys urticae* data in Section 3.5.1 are used. In addition in case only a number of the parameters are included in the (init_list) element then the other parameters are initialized according to the 'rstan' library initialization rules (Stan Development Team 2021*a*).

Optionally user can also apply other argument choices as defined and supported by the *stan* function in the 'rstan' library in R platform.

4.2.5 Output

The *drin_hmc* function returns a list containing three distinct objects. To begin, the estimation element (est) is a Stanfit object of type s4 that retains all of its properties (Stan Development Team 2021*b*). The 'est' object also includes the (dev) parameter, which represents the current model's deviance, as well as the (ypred) variable, which contains the current model's posterior predictive values based on the xpred imposed values. Second, the Stanfit summary statistics of the fundamental model parameters are denoted as (summary) argument. Third, includes Stanfit's overall diagnostics report (diagnostics) argument, which includes the results of the *check_hmc_diagnostics* function of the 'rstan' package (Stan Development Team 2021*b*), which contains several diagnostic tools with percentages of iterations. 1. that resulted in a divergence, 2. that exhausted the maximum treedepth, and 3. the energy Bayesian fraction of missing data (E-BFMI) (Betancourt 2017). Additionally, the (C_data), the (mtype) and the (lik) elements are attached in the final output so that the user can recall them at his leisure. Finally, based on the predictor $xpred$ values, a 95% credible intervals posterior predictive plot is created alongside the original data points using the *drin_popp* function. This plot can be used to test the model's applicability to the current data set (when $xpred$ values are chosen close to the working data) or to predict the developmental rates at different temperatures given the imposed model.

4.3 Examples

In this section, examples for using the 'DRIN' package are provided so that the user not only becomes familiar with the call of the basic function, but also understands the structure and content of the output without further testing, allowing him to make better use of it.

4.3.1 *Neoseiulus californicus* example

The dataset 1 consists of 212 cases of mite *Neoseiulus californicus* that have reached adult stage until the study ended. Developmental rates derived by calculating the reciprocal of the observed days passed until the mite's state change and have minimum 0.028 at 12.5^oC and maximum 0.333 at 32.5^oC.

4.3.1.1 Experimental design

The four ecological models are used both assuming the Gaussian and the Inverse gamma distributions for the data.

Before arriving at calling the generic function, the following steps must be completed.

- Install and load the 'DRIN' package: The package is deposit in GitHub repository which requires to install and load the 'devtools' package first.
- Build the data list and select the arguments to use: In the initial step the current data list should contain the responses under the name "y" and the corresponding temperatures under the name "x". Optionally, one can add the other arguments mentioned on sections 4.2.1, 4.2.2, 4.2.3 and 4.2.4.
- Call the *drin_hmc* function Call *drin_hmc* function using the selected arguments and save the output list in a desired name.
- Call the *drin_popp* function Call *drin_popp* function to produce a posterior predictive plot along with the working data. The arguments needed are the data to plot (data), the Rstan object that contains the predicted values (est), the model type (mtype) and the likelihood distribution of the model (lik). The output of *drin_hmc* function contains all the required arguments to call the *drin_popp* function.

```
mainDir = "c:\Main␣directory␣on␣local␣drive"
subDir = "Current␣directory␣address␣here"
setwd(file.path(mainDir, subDir))
#If not pre-installed first install devtools library
library(devtools) #Load devtools to access GitHub.
install_github("mkondakis/DRIN") #Install DRIN library
library(DRIN)
#Create data list with the minimum elements:
data = list(y = y, x=temp)
data #print data list
```

```
#Call the drin_hmc function to sample using HMC via STAN
modbieri = drin_hmc(data,mtype="bieri",+
lik="gauss",prior = list(a="exponential(0.001)"))
drin_popp(data=modbieri$data,est=modbieri$est,+
mtype=modbieri$mtype,lik=modbieri$lik)
```

Listing 4.1: R output of code at line 10

```
$y
  [1] 0.029 0.034 0.032 0.037 0.034 0.048 0.034 0.037 0.037 0.032 0.028 0.038 0.030 0.038
 [15] 0.040 0.034 0.034 0.034 0.083 0.062 0.062 0.067 0.059 0.062 0.062 0.062 0.071 0.067
 [29] 0.083 0.077 0.059 0.062 0.071 0.062 0.059 0.077 0.062 0.100 0.091 0.091 0.091 0.100
 [43] 0.083 0.083 0.091 0.083 0.100 0.091 0.083 0.100 0.083 0.100 0.091 0.091 0.083 0.091
 [57] 0.083 0.083 0.083 0.091 0.083 0.083 0.100 0.100 0.100 0.167 0.143 0.125 0.143 0.125
 [71] 0.111 0.143 0.167 0.143 0.182 0.125 0.125 0.125 0.143 0.143 0.125 0.167 0.143 0.143
 [85] 0.143 0.125 0.200 0.125 0.167 0.167 0.167 0.167 0.167 0.200 0.167 0.167 0.167 0.167
 [99] 0.200 0.143 0.167 0.200 0.167 0.167 0.200 0.167 0.200 0.167 0.167 0.222 0.222 0.222
[113] 0.222 0.250 0.222 0.250 0.222 0.200 0.250 0.222 0.200 0.222 0.200 0.200 0.222 0.222
[127] 0.200 0.200 0.200 0.200 0.222 0.200 0.200 0.222 0.200 0.250 0.286 0.222 0.250 0.250
[141] 0.250 0.222 0.222 0.222 0.250 0.250 0.250 0.222 0.222 0.250 0.286 0.286 0.286 0.286
[155] 0.250 0.250 0.250 0.250 0.250 0.286 0.286 0.250 0.250 0.250 0.250 0.250 0.250 0.250
[169] 0.250 0.250 0.250 0.250 0.286 0.286 0.250 0.286 0.286 0.286 0.286 0.286 0.286 0.333
[183] 0.286 0.333 0.286 0.286 0.333 0.333 0.333 0.286 0.250 0.286 0.286 0.333 0.286 0.222
[197] 0.250 0.250 0.250 0.250 0.222 0.250 0.286 0.200 0.286 0.250 0.222 0.286 0.250 0.286
[211] 0.250 0.250

$x
  [1] 12.5 12.5 12.5 12.5 12.5 12.5 12.5 12.5 12.5 12.5 12.5 12.5 12.5 12.5 12.5 12.5 12.5
 [18] 12.5 15.0 15.0 15.0 15.0 15.0 15.0 15.0 15.0 15.0 15.0 15.0 15.0 15.0 15.0 15.0 15.0
 [35] 15.0 15.0 15.0 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5
 [52] 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 20.0 20.0 20.0
 [69] 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0
 [86] 20.0 20.0 20.0 22.5 22.5 22.5 22.5 22.5 22.5 22.5 22.5 22.5 22.5 22.5 22.5 22.5 22.5
[103] 22.5 22.5 22.5 22.5 22.5 22.5 22.5 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0
[120] 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 27.5
[137] 27.5 27.5 27.5 27.5 27.5 27.5 27.5 27.5 27.5 27.5 27.5 27.5 27.5 27.5 30.0 30.0 30.0
[154] 30.0 30.0 30.0 30.0 30.0 30.0 30.0 30.0 30.0 30.0 30.0 30.0 30.0 30.0 30.0 30.0 30.0
[171] 30.0 30.0 32.5 32.5 32.5 32.5 32.5 32.5 32.5 32.5 32.5 32.5 32.5 32.5 32.5 32.5 32.5
[188] 32.5 32.5 32.5 32.5 32.5 32.5 32.5 32.5 35.0 35.0 35.0 35.0 35.0 35.0 35.0 35.0 35.0
[205] 35.0 35.0 35.0 35.0 35.0 35.0 35.0 35.0
```

4.3.1.2 Estimation of the parameters

The output of the code line 10 is a list containing several parts shown in the listings below.

Listing 4.2: R output

```
$est
Inference for Stan model: 4877488a7f849423429f139bb5d5baa5.
3 chains, each with iter=11000; warmup=10000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=3000.

                 mean se_mean    sd      2.5%       25%
50%       75%    97.5% n_eff Rhat
a                0.01    0.00  0.00      0.01      0.01      0.01
0.01      0.01     814 1.00
b                2.19    0.12  2.22      1.52      1.71      1.85
2.12      4.59     363 1.00
tmin            10.20    0.01  0.27      9.63     10.02     10.21
10.38     10.71    882 1.00
tmax            38.77    0.04  0.93     36.59     38.18     38.82
39.36     40.46    692 1.00
sigmasq          0.00    0.00  0.00      0.00      0.00      0.00
0.00      0.00    1676 1.00
sigma            0.02    0.00  0.00      0.02      0.02      0.02
0.02      0.02    1681 1.00

Samples were drawn using NUTS(diag_e) at Tue Aug 24 21:19:41 2021.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).

$summary
                mean      se_mean sd              2.50%           25%             50%
a               0.01      0.00    0.00            0.01            0.01            0.01
b               2.19      0.12    2.22            1.52            1.71            1.85
tmin            10.20     0.01    0.27            9.63            10.02           10.21
tmax            38.77     0.04    0.93            36.59           38.18           38.82
sigma           0.02      0.00    0.00            0.02            0.02            0.02
dev             -1092.780.14      3.60            -1097.39        -1095.41        -1093.59
tdmax           32.74     0.01    0.28            32.34           32.56           32.70
xmin[1]         10.20     0.01    0.27            9.63            10.02           10.21
xmax[1]         37.22     0.02    0.58            35.91           36.85           37.24
lp__            733.00    0.07    1.73            728.58          732.16          733.39
                75%               97.50%          n_eff           Rhat
a               0.01              0.01            814.17          1.00
b               2.12              4.59            362.81          1.00
tmin            10.38             10.71           882.28          1.00
tmax            39.36             40.46           692.23          1.00
sigma           0.02              0.02            1680.81         1.00
dev             -1091.07          -1083.61        670.59          1.00
tdmax           32.87             33.52           485.13          1.01
xmin[1]         10.38             10.71           882.54          1.00
xmax[1]         37.58             38.31           721.03          1.00
lp__            734.34            735.32          648.09          1.00
```

The first listing, 4.3.1.2, contains the (est) part, which is the Stanfit s4' object that retains all of its properties and can be used in the same way as an 'rstan' output. Second, a summary statistics (summary) of the parameters of interest is provided, which includes not only the mean, MCMC error, standard deviation, and quartiles, but also the effective sample size and the Gelman-Rubin convergence diagnostic R.

Listing 4.3: R output

```
$diagnostics
[1] "0 of 3000 iterations ended with a divergence."
[2] "0 of 3000 iterations saturated the maximum tree depth of 11."
[3] "E-BFMI indicated no pathological behavior."

$data
$data$y
 [1]  0.03  0.03  0.03  0.04  0.03  0.05  0.03
[8]   0.04  0.04  0.03  0.03  0.04  0.03  0.04
[15]  0.04  0.03  0.03  0.03  0.08  0.06  0.06
[22]  0.07  0.06  0.06  0.06  0.06  0.07  0.07
[29]  0.08  0.08  0.06  0.06  0.07  0.06  0.06
[36]  0.08  0.06  0.10  0.09  0.09  0.09  0.10
[43]  0.08  0.08  0.09  0.08  0.10  0.09  0.08
[50]  0.10  0.08  0.10  0.09  0.09  0.08  0.09
[57]  0.08  0.08  0.08  0.09  0.08  0.08  0.10
[64]  0.10  0.10  0.17  0.14  0.13  0.14  0.13
[71]  0.11  0.14  0.17  0.14  0.18  0.13  0.13
[78]  0.13  0.14  0.14  0.13  0.17  0.14  0.14
[85]  0.14  0.13  0.20  0.13  0.17  0.17  0.17
[92]  0.17  0.17  0.20  0.17  0.17  0.17  0.17
[99]  0.20  0.14  0.17  0.20  0.17  0.17  0.20
[106] 0.17  0.20  0.17  0.17  0.22  0.22  0.22
[113] 0.22  0.25  0.22  0.25  0.22  0.20  0.25
[120] 0.22  0.20  0.22  0.20  0.20  0.22  0.22
[127] 0.20  0.20  0.20  0.20  0.22  0.20  0.20
[134] 0.22  0.20  0.25  0.29  0.22  0.25  0.25
[141] 0.25  0.22  0.22  0.22  0.25  0.25  0.25
[148] 0.22  0.22  0.25  0.29  0.29  0.29  0.29
[155] 0.25  0.25  0.25  0.25  0.25  0.29  0.29
[162] 0.25  0.25  0.25  0.25  0.25  0.25  0.25
[169] 0.25  0.25  0.25  0.25  0.29  0.29  0.25
[176] 0.29  0.29  0.29  0.29  0.29  0.29  0.33
[183] 0.29  0.33  0.29  0.29  0.33  0.33  0.33
[190] 0.29  0.25  0.29  0.29  0.33  0.29  0.22
[197] 0.25  0.25  0.25  0.25  0.22  0.25  0.29
[204] 0.20  0.29  0.25  0.22  0.29  0.25  0.29
[211] 0.250 0.250
```

Moreover the second listing 4.3.1.2 contains the diagnostics report (diagnostics) argument, which includes 1. that resulted in a divergence, 2. that exhausted the maximum

treedepth, and 3. the energy Bayesian fraction of missing data (E-BFMI) (Betancourt 2017). Additionally the working data (x: predictor values, y: response values) are included so that the user can recall and reuse them in supporting functions such as the *drin_popp'function*.

Listing 4.4: R output

```
$data$x
  [1] 12.5 12.5 12.5 12.5 12.5 12.5 12.5 12.5 12.5 12.5 12.5 12.5 12.5 12.5 12.5 12.5 12.5 12.5
 [18] 12.5 15.0 15.0 15.0 15.0 15.0 15.0 15.0 15.0 15.0 15.0 15.0 15.0 15.0 15.0 15.0 15.0 15.0
 [35] 15.0 15.0 15.0 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5
 [52] 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 17.5 20.0 20.0 20.0
 [69] 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0
 [86] 20.0 20.0 20.0 22.5 22.5 22.5 22.5 22.5 22.5 22.5 22.5 22.5 22.5 22.5 22.5 22.5 22.5 22.5
[103] 22.5 22.5 22.5 22.5 22.5 22.5 22.5 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0
[120] 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 25.0 27.5
[137] 27.5 27.5 27.5 27.5 27.5 27.5 27.5 27.5 27.5 27.5 27.5 27.5 27.5 27.5 30.0 30.0 30.0
[154] 30.0 30.0 30.0 30.0 30.0 30.0 30.0 30.0 30.0 30.0 30.0 30.0 30.0 30.0 30.0 30.0 30.0 30.0
[171] 30.0 30.0 32.5 32.5 32.5 32.5 32.5 32.5 32.5 32.5 32.5 32.5 32.5 32.5 32.5 32.5 32.5 32.5
[188] 32.5 32.5 32.5 32.5 32.5 32.5 32.5 32.5 35.0 35.0 35.0 35.0 35.0 35.0 35.0 35.0 35.0 35.0
[205] 35.0 35.0 35.0 35.0 35.0 35.0 35.0 35.0

$data$N
[1] 212

$data$status
  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [44] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [87] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[130] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[173] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

$data$t
[1] 10

$data$NP
[1] 45

$data$xpred
 [1] 12.50 13.01 13.52 14.03 14.55 15.06 15.57 16.08 16.59
[10] 17.10 17.61 18.13 18.64 19.15 19.66 20.17 20.68 21.19
[19] 21.70 22.22 22.73 23.24 23.75 24.26 24.77 25.28 25.80
[28] 26.31 26.82 27.33 27.84 28.35 28.86 29.38 29.89 30.40
[37] 30.91 31.42 31.93 32.44 32.95 33.47 33.98 34.49 35.00



$mtype
[1] "bieri"

$lik
[1] "gauss"
```

Furthermore, the third listing /refc4_lstin4 contains the data size (N), predictor values (xpred), prediction size (NP), model type (mtype), and distribution of the data considered (lik), all of which are available to the user for further use.

4.3.1.3 Graphical representations

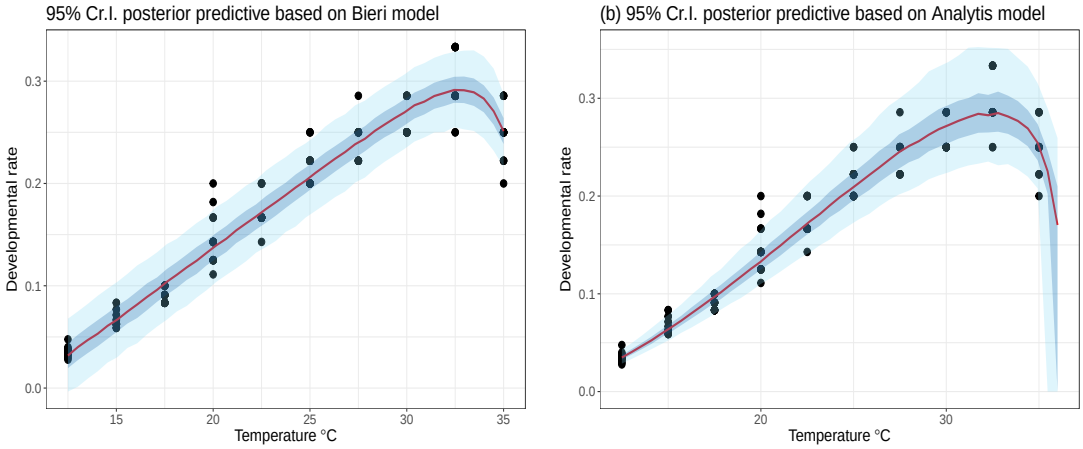

Figure 4.1 Posterior predictive plots for (a) Gaussian and Bieri and (b) Inv. Gamma and Analytis models for *Neoseiulus californicus* mites data calling the *drin_popp* function.

The *drin_popp* function is used in the code line 15. The input arguments used are the (data), (est), (mtype) and (lik) output elements of the 'modbieri' object created by the *drin_hmc* function called in line 13 in the aforementioned code listing. Figure 4.1(a) is the output of the code line 15. We find that the model based on Bieri and the Gaussian model fits the working data well, while the model results based on Analytis and Inverse Gamma is variance adaptive as shown in 4.1(b). Parameter estimations can be used by ecologists to infer the developmental dynamics of *Neoseiulus californicus* under relative conditions, while specialists can combine those figures with other biological features and proceed with mite management actions where necessary.

4.4 Chapter summary

In this Chapter, we present a modern R-package that combines the various nonlinear ecological functions and data distributions described in Chapter 3 into a single user-friendly function for estimating model parameters of interest and inference for arthropod developmental rates. Furthermore, the proposed procedure not only provides default reliable options for the parameters prior distributions and initial values, but also allows the user to change them at his or her discretion. Besides that, the model code used, which makes use of STAN software capabilities, can address not only the difficulties of the restricted domains of some parameters and ecological functions, but also the excesses zeros in the observed data. Finally, it provides posterior predictive and posterior deviance estimates, which can be used to check and compare models. More options for ecological functions and data distributions can be added in future work.

Chapter 5

Modelling Predator-prey systems

5.1 Introduction

The study of predator-prey population interactions is critical in ecology for understanding how to describe and predict population performance, as well as how to manage invasive species. Devastating species can seriously damage edible commodities (Paini et al. 2016) or interfere with other species by preying on native species or out-competing native species for food or other resources (Marras et al. 2015). They can also cause or carry vector-borne diseases (Medlock and Leach 2015), harm the food chain (Engeman et al. 2010) and the ecosystem in a variety of ways (Pejchar and Mooney 2009).

The dynamics of predator-prey interactions can indicate complex behavior in many situations. The type of competition relating to feeding rate can be used to classify these interactions. The exploitation competition, which is the amount of remaining resource that has already been exploited by others and is thus inaccessible, the interference competition, which occurs when foraging individuals interact directly, and the mutual interference, which involves individuals of the same species.

Defused by analysis in (Papanikolaou et al. 2016b), in which the model schemes account for predator-predator interactions, and in order to improve model predictability, we incorporate a stochastic process into estimating the probability of prey consumption. When the volatility of this stochastic approach becomes negligible, it is reduced to the deterministic approach, in which the probability of prey consumption is explicitly linked to the rate of prey consumption. Furthermore, we employ and compare both HMC and VBI techniques to obtain posterior samples of target parameter distributions. Real datasets from the predation of the *Propylea quatuordecimpunctata* beetle over fabae scopoli aphids (Papanikolaou et al. 2016b) are analyzed to evaluate and compare the performance of the deterministic approach to the stochastic one. The findings of Chapter 5 would be submitted for review in the Autumn of 2021.

The rest of this chapter is structured as follows. Section 5.2 describes the elements that compose the model class that we incorporate in both the deterministic and stochastic approaches. Section 5.3 investigates four separate real-life examples without and with predator interference, and Section 5.3.3 reports our findings, and the chapter con-

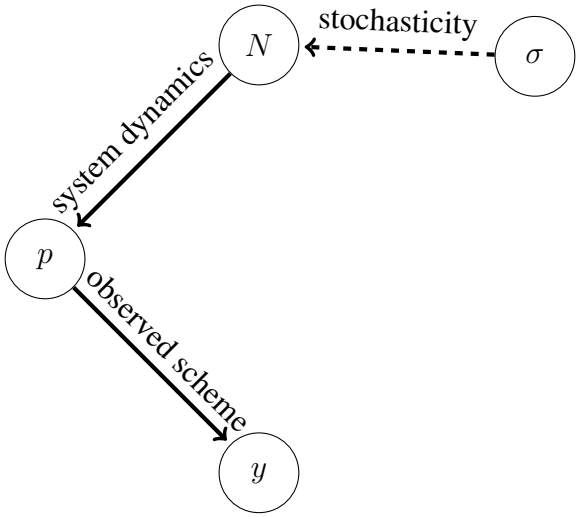

Figure 5.1 Graphical representation of the structure of the suggested stochastic model.

cludes with discussion.

5.2 Modelling class approach

The modelling class used in this work is divided into four components, as shown in Fig. (5.1). The sampling distribution of the data y is the observational component that generates the counts of prey consumed at specific time intervals. The Binomial distribution is chosen, with p denoting the probability of being consumed in a given time interval. The probability p is related to the abundance of the prey density N, which can be described by ode ecological models in real time. This connection can be made explicitly via the ratio of prey consumed in a given time interval, which will be referred to as the deterministic approach from now on, or implicitly through a stochastic process. which will be referred to as the stochastic approach from now on. The stochastic process is the OU process, the value of which is linked to the probability p, the mean of which is linked to the ratio of prey consumed, and the volatility of which is linked to the diffusion term σ. When σ is zero, the stochastic approach is reduced to the deterministic one.

5.2.1 Binomial model

To begin, we consider a suitable distribution for producing the data denoted by y_{t_i}. Data, y_{t_i} represent the number of prey consumed at time instant t_i. In addition, y_{t_i} are considered independent and identically distributed response variables, where $i = 1, \ldots, n$ are the n repeated measures. Because of their simplicity and nice properties, the Normal and Log-normal are commonly used in this context. Nonetheless, the data observed are discrete counts, and it is not appropriate to generate them using continuous distributions Fernández and Steel (1998). As a result, we employ the Binomial distribution

with parameters N_0 and p_{t_i} as follows:

$$y_{t_i} \sim Binom(N_0, p_{t_i}), \tag{5.1}$$

where N_0 represents the total number of prey at the start of the study and p_{t_i} denotes the probability that the prey subject has been consumed by time t_i.

The probability p_{t_i} varies over time and is linked to the ratio of prey consumed to prey abundance (instantaneous ratio of consumed prey) during the study period. To this end, p_{t_i} can be derived from the current prey density either directly in a deterministic way or indirectly through stochasticity. The latter is use full so as to compensate for noise not only in the data collected, a predominant situation in ecology, but also due to possible model mis-specification, its common practise to introduce complexity in the classical Binomial model 5.1 using stochastic processes. (Malesios et al. 2017, Zhu et al. 2017).

The p_{t_i} in this case is generated by the ratio of the prey consumed at time t_i (instantaneous ratio of consumed prey density) as shown in the following formula:

$$p_{t_i} = \frac{N_0 - N_{t_i}}{N_0}, \tag{5.2}$$

where N_{t_i} represents the instantaneous prey population density at time t_i. The prey density N_{t_i} evolves over time and its value variation can be explained by ecological models that involves differential equations.

5.2.1.1 Ecological models

Differential equations have been used to describe predator-prey dynamics in a variety of settings (Arditi and Ginzburg 2012, Berryman 1992). An early example is the conventional Lotka and Volttera couple of differential equations (Lotka 1925, Volterra 1926), where predator and prey population rate change is exclusively connected with the number of individuals in the system and the predator-prey interaction. In the predator absence the prey reproduces exponentially in this dynamic system depending on its intrinsic growth rate r, assuming an unlimited supply of food. In the presence of predation on the prey, the rate of prey density decreases or the rate the predator density increases is proportional to the rate at which the predators and prey meet.

The formulation Lotka-Volttera method is fundamental in ecology, but it fails to explain saturation not only in terms of prey consumption rate when prey populations burst, but also in terms of resources depletion when predators are absent from the system and prey have reached their maximum capacity. In order to improve on the former weakness, Holling (Holling 1959*a*,*b*) introduced the concept of functional response $f(N)$, which denotes the change in the number of prey captured by the predator per

unit of time. Specifically, the functional response $f(N)$ in Lotka-Volttera method is equal to $f(N) = r \cdot N$ which is a linear increasing function of prey abundance only and does not involve predator behavioral characteristics. Several functional response forms have been proposed, but Holling's work (Holling 1959*a*,*b*), which imports certain predator behavior types, is a cornerstone and is broadly used in the literature. type I assumes a continuous linear response as the prey population grows to a certain size, whereas other types assume that the consumption rate pattern does not always increase but begins to stabilize once a certain level of prey density is reached. types II and III, in particular, exert a decelerating and sigmoid response, respectively. Rosenzweig and MacArthur (Rosenzweig and MacArthur 1963, Rosenzweig 1969, 1977) studied similar models combined Holling functional response with a logistic growth in the absence of a predator.

We focus on prey population change due to predation in this study and assume that prey and predator density change is negligible as a result of reproduction or migration. This situation arises, for instance, when studying system dynamics in closed patches under controlled laboratory conditions and predator or prey reproduction begins after the study is completed. In this manner, we investigate prey consumption rate when predator density is held constant at certain concentrations. Our reference model is based on the commonly used Holling type II functional response.

5.2.1.1.1 Holling type II. The type II functional response Holling (1959*b*) assumes a slowing consumption rate as a result of the predator's limited food processing capacity, which leads to saturation. A type II functional response of a hyperbolic shape characterizes the Holling disc equation, which is described by the formula:

$$\frac{dN}{dt} = -\frac{aNP}{1 + aT_hN} \tag{5.3}$$

where $N = N_t$ and $P = P_t$ denote the changing prey and predator densities over time and $\frac{dN}{dt}$ represents the rate of change of prey density. Moreover, a represents the predator's attack rate or rate of seeking efficiency, which describes the prey mortality per predator at spatially homogeneous low prey densities, and P represents the predator density. Finally, T_h denotes a predator's handling time in hours, which represents the time spent pursuing, subduing, eating, and digesting its prey.

Functional responses can take the form of either pure prey dependence $g(N)$, which is a function of the absolute population of preys N, or ratio dependence $g(N/P)$, which is a function of the prey ration of each of P predators. The former is more appropriate in cases characterized by slow predator growth and low densities of prey that are spatially distributed in a uniform manner throughout the study area. The latter case, which is outside the scope of our study, is more suitable in situations of higher predator densities

or/and mixed predator communities, while using a longer demographic time scale for population dynamics (Arditi and Ginzburg 2012).

Additionally, theoretical and empirical studies (Skalski and Gilliam 2001, Kratina et al. 2009, Arditi and Ginzburg 2012) emerge the importance of incorporating behavioral interactions amongst foraging predators. Specific modelling schemes that account for interference competition have been developed, using either phenomenological approaches that describe the empirical relationship between functional response and predator density or mechanistic approaches that incorporate variables with clear interpretation approaches (Hassell and Varley 1969, Beddington 1975, DeAngelis et al. 1975, Crowley and Martin 1989, Arditi and Ginzburg 2012). To account for interference effects in the predation process, the functional response in this context is both prey and predator dependent $f(N, P)$. To compensate for interactions betweeen predators, the Beddington-DeAngelis, Crowely-Martin, and Hassell-Varley odes have been used for comparing purposes as in (Papanikolaou et al. 2016b). Moreover, We use the Bayesian paradigm and the HMC method, as well as the ADVI-meanfield and ADVI-fullrank alternative computational methods. The STAN software platform (Stan Development Team 2021*b*) is used not only to perform statistical inference for the parameters of interest, but also to compare the various computational methods in such models. Aside from that, the STAN software (Stan Development Team 2021*b*) provides built-in mechanisms called "rk45" and "bdf" to approximate the solution of the corresponding ode using a fourth and fifth order Runge-Kutta method for non-stiff systems (Dormand and Prince 1980) and backward-differentiation formula implementation for stiff systems (Cohen et al. 1996), respectively.

In order to account for predator interference, we employ the following ode methods:

5.2.1.1.2 Beddington-DeAngelis ode. The common Beddington-DeAngelis ecological mechanistic method developed separately by Beddington (Beddington 1975) and (DeAngelis et al. 1975) extent the purely-dependent Holling type II method to incorporate predators density. It is described by the formula:

$$\frac{dN}{dt} = -\frac{aNP}{1 + aT_hN + \beta t_w(P-1)} \tag{5.4}$$

where β denotes the rate that a single predator encounters to the other predators which occur with period t_w. Because the product βt_w which describes the interference magnitude contains two unknown parameters that are treated as random variables in the Bayesian context, we use only one parameter in our analysis, $c = \beta t_w$.

5.2.1.1.3 Crowely-Martin ode. Crowely-Martin (Crowley and Martin 1989) developed another mechanistically method based on the Beddington DeAngelis method in

5.4 and is described by the formula:

$$\frac{dN}{dt} = -\frac{aNP}{1 + aT_hN + \beta t_w(P-1) + \alpha T_h \beta t_w N(P-1)} \tag{5.5}$$

In our analysis, we use the parameter $c = \beta t_w$, as in the Beddington DeAngelis ode case (5.4).

5.2.1.1.4 Hassell-Varley ode. Hassell and Varley (Hassell and Varley 1969) developed a phenomenological ode method, which Sutherland modified in (Sutherland 1983) and is described by the formula:

$$\frac{dN}{dt} = -\frac{aNP^{-m}}{1 + aT_hNP^{-m}} \tag{5.6}$$

where m denotes the magnitude of predator-prey interference competition. In case $m = 0$, the equation (5.6) becomes the Holling type II ode.

In case $P = 1$, the previous three models in (5.4, 5.5, 5.6) are equal to the Holling type II model in (5.3).

5.2.2 Stochastic models

In this context, we suggest involving a stochastic process in deriving the probability of the Binomial distribution in (5.1) not only to absorb potential model mis-specification, but also to account for noise from the nature of the ecological data, allowing for more accurate parameter estimations. The Ornstein–Uhlenbeck OU process κ_s can be thought of as the time-continuous AR(1) analog and is used to connect the probability p_{t_i} in (5.1) to the piece-wise mean of the process as follows:

$$\kappa_t = log\left(\frac{p_t}{1 - p_t}\right) \tag{5.7}$$

$$d\kappa_s = \phi\left(\mu_t - \kappa_s\right) ds + \sigma dB_s \tag{5.8}$$

$$\mu_t = logit\left(\frac{N_0 - N_t}{N_0}\right), \tag{5.9}$$

where B_s is standard Brownian motion, σ is the diffusion term, ϕ is the speed of reversion to the mean and μ_t a piece-wise constant function determined by the logarithm of the odds of the ratio of the consumed prey up to time t. The instantaneous process κ_t is the OU process that evolves around mean μ_t, while its transition density from time point t to t+1 is given explicitly by:

$$k_{t+1|t} \sim N\left(\mu_{t+1} + \left(d\kappa_t - \mu_{t+1}\right) exp(-\phi), \psi\right) \tag{5.10}$$

Table 5.1 Priors of the parameters of the four ecological ode model.

Models	Parameters	Priors
Holling disc	α	Exp(10^{-1})
	T_h	Exp(10^{-1})
Crowely Martin	α	Exp(10^{-1})
	T_h	Exp(10^{-1})
	c	Exp(10^{-1})
Beddington DeAngelis	α	Exp(10^{-1})
	T_h	Exp(10^{-1})
	c	Exp(10^{-1})
Hassel Varley	α	Exp(10^{-1})
	T_h	Exp(10^{-1})
	m	Uniform$(-1, 1)$
(O-U) process parameters		
Reversion speed	ϕ	$\Gamma\left(10^{-1}, 10^{-1}\right)$
Diffusion term	σ^2	$\Gamma\left(10^{-1}, 10^{-1}\right)$

where $\psi = \frac{\sigma^2}{2\phi}\left(1 - exp(-2\phi)\right)$ is the variance of the Gaussian density in (5.10). In the absence of a diffusion term, the stochastic process converges to its mean μ_t which now equals the instantaneous ratio of consumed prey density as equations (5.7 and 5.9) become equal, resulting in the deterministic case in (5.2). The variance of the transition distribution ψ becomes zero in this case.

5.2.3 Priors

In the ode ecological models under study (5.3, 5.4, 5.5, 5.6) we use the exponential distribution with parameter 10^{-1} as the prior distribution for the parameters of interest as shown in Table 5.1. The priors used for the same ode are identical, whereas the stochastic models includes two extra parameters σ and ϕ.

5.3 Real-life application

Predaceous coccinellidae are insect pest predators over aphids and coccids Dixon et al. (2009). Aphidophagous species, such as *Propylea quatuordecimpunctata*, are effective at suppressing aphid populations and exhibit aggregation to their prey (Hodek and Honêk 2013, Schellhorn and Andow 2005), making them popular in biological control, especially in the short term Obrycki et al. (2009). Furthermore, these data only include larvae, which typically live their entire lives within a patch, as opposed to adults, who are distinguished by their ability to fly (Kindlmann and Dixon 2001, Dostalkova et al. 2002). As larvae socialize, frequent encounters between individuals may affect their

foraging success.

5.3.1 Data specifications

The data used provided in (Papanikolaou et al. 2016b) comprises of two species involving as predator and prey respectively. To begin the fourteen-spotted ladybird predator beetle *Propylea quatuordecimpunctata* L. (Coleoptera: Coccinellidae) (Day et al. 1994, Hodek and Honêk 2013) is found throughout Europe. On the other hand the black bean aphids (*Aphis fabae Scopoli*) can influence the growth rates of important plants for food production (HUREJ and WERF 1993, Basedow et al. 2006). *Propylea quatuordecimpunctata* is an aphid-suppressing predator that forms aggregations with its prey (Hodek and Honêk 2013, Schellhorn and Andow 2005) . The dataset used include predator larvae aged 0.5 to 1.5 days, which are established aphidophagous arthropods (Papanikolaou et al. 2016a). Also included as prey are immature (1^st, 2^nd, 3^nd and 4^th instars and adults) black bean aphids (*Aphis fabae Scopoli*)) aged 3 to 3.5 days. In the presence of one, two, three, and four predators, four prey groups were tested. The prey densities evaluated were 5, 10, 15, 20, and 25 aphid nymphs for individual predators, 10, 20, 30, 40, and 50 nymphs for two predators, 15, 30, 45, 60, and 75 nymphs when predator density was three larvae, and 20, 40, 60, 80, and 100 nymphs when predator density was four larvae. The exposure time was 6 hours. Higher predator concentrations are not explored since this might result in larvae crowding and/or cannibalism, which is typical in laboratory-reared coccinellids, particularly at low prey densities (Papanikolaou et al. 2016a,b).

5.3.2 Estimation of the parameters

The computational methods used are the HMC method performed using 3 chains, each of 30000 iterations with 20000 warmup samples, while for ADVI ADVI methods we used 10000 iterations and a tolerance level of 0.01. Posterior estimates of the mean and the 95% Cr.I. for the deterministic and stochastic models studied in section 5.2 approach are shown together in Table 5.2 so that the results can be compared not only between the two approaches but also among the different predator density and ecological ode methods.

Table 5.2 Posterior means and 95% Cr.I. for deterministic and stochastic models. The HMC computation method is used.

		Deterministic models				Stochastic models			
	ODE	α	T_h	c	m	α	T_h	c	m
P=1	H2†	0.460	0.215			0.726	0.295		
		0.345-0.603	0.148-0.276			0.511-0.965	0.214-0.383		
P=2	H2†	0.286	0.294			0.37	0.313		
		0.23-0.351	0.252-0.334			0.268-0.515	0.236-0.393		
	BD‡	0.304	0.282	0.151		0.306	0.227	0.163	
		0.255 -0.361	0.242-0.319	0.006-0.446		0.252-0.367	0.133-0.308	0.005-0.531	
	CM§	0.302	0.250	0.142		0.306	0.203	0.147	
		0.254-0.358	0.179-0.306	0.005-0.423		0.252-0.366	0.106-0.293	0.005-0.49	
	HV§§	0.302	0.141		0.815	0.305	0.114		0.809
		0.254-0.359	0.121-0.159		0.494-0.992	0.252-0.365	0.067-0.155		0.422-0.994
P=3	H2†	0.332	0.339			0.393	0.298		
		0.278-0.394	0.315-0.362			0.274-0.642	0.214-0.378		
	BD‡	0.349	0.333	0.060		0.346	0.242	0.089	
		0.303-0.403	0.31-0.354	0.002-0.174		0.292-0.404	0.162-0.316	0.003-0.279	
	CM§	0.348	0.301	0.056		0.345	0.212	0.081	
		0.302-0.4	0.24-0.345	0.002-0.165		0.292-0.403	0.129-0.296	0.003-0.257	
	HV§§	0.348	0.111		0.905	0.345	0.081		0.870
		0.302-0.401	0.104-0.118		0.738-0.996	0.292-0.404	0.055-0.105		0.622-0.995
P=4	H2†	0.186	0.345			0.249	0.339		
		0.157-0.218	0.317-0.371			0.178-0.349	0.251-0.425		
	BD‡	0.194	0.337	0.038		0.196	0.260	0.043	
		0.17-0.223	0.311-0.361	0.002-0.109		0.168-0.226	0.171-0.34	0.001-0.147	
	CM§	0.194	0.307	0.035		0.195	0.237	0.039	
		0.169-0.222	0.246-0.351	0.001-0.104		0.167-0.226	0.14-0.326	0.001-0.136	
	HV§§	0.194	0.084		0.928	0.195	0.066		0.922
		0.17-0.222	0.078-0.09		0.803-0.997	0.167-0.225	0.043-0.085		0.754-0.998

† Holling type II.
‡ Beddington de Angelis.
§ Crowely Martin.
§§ Hassell Varley.

Table 5.3 Model selection criteria for the four models applied in the deterministic approach.

DETERMINISTIC CASE

2 Predators				
	Holling II	Beddington	Crowely	Hassell
AIC	**359.98**	362.04	361.94	361.96
BIC	**365.72**	369.68	369.59	369.61
DIC	356.04	356.17	**355.89**	355.94
LooIC	362.56	361.38	**361.09**	361.18
WAIC	362.44	361.34	**361.04**	361.14
3 Predators				
	Holling II	Beddington	Crowely	Hassell
AIC	**535.78**	537.81	537.74	537.76
BIC	**541.52**	545.46	545.39	545.4
DIC	531.75	531.88	**531.7**	531.7
LooIC	545.19	543.36	**543.1**	543.24
WAIC	545.08	543.32	**543.06**	543.22
4 Predators				
	Holling II	Beddington	Crowely	Hassell
AIC	**662.01**	664.03	663.96	663.97
BIC	**667.75**	671.67	671.61	671.62
DIC	658.1	658.09	**657.89**	657.92
LooIC	673.11	671.78	**671.62**	671.64
WAIC	673.07	671.68	**671.55**	671.59

5.3.2.1 Predator interaction effect

In the case of a single predator (P=1), all models are equivalent to the Holling type II model, and the predator's mean handling time falls within the 95% Cr.I. intervals (0.15h, 0.28h) and (0.21h, 0.38h) for the deterministic and the stochastic case respectively. The estimated handling times of Hassell Varley decreased as predator density increased in comparison to the other models, but there is no statistically significant deviation in any of the four ecological models tested. This finding suggests that predators' per capita feeding rate and predator density are roughly independent. When the number of predators is less than four, there are no significant differences in attack rates, as shown in Table 5.2. However, the presence of four predators results in a statistically significant decrease in attack rates in all models, both deterministic and stochastic. The 95% Cr.I. of magnitude of interference represented by variables c and m excludes the value zero, implying that interference between predators is not negligible. Furthermore, estimates of variables c and m showed no significant differences as the number of predators increased, indicating that the magnitude of interference has no effect on the fluctuation of predator density.

5.3.2.2 Model comparison

In all ode methods studied, both model approaches -deterministic and stochastic- provide meaningful parameter of interest estimates.

In addition figures 5.2 and 5.3 reveal that the 95% posterior predictive curves exceeds almost all of the bulk data values. The posterior predictive probabilities demonstrate that the investigated models fit the data properly. Furthermore, stochastic models outperform deterministic models in terms of prediction efficiency. This discovery leads to the conclusion that the OU process can successfully handle noise in data that can arise from a variety of sources, such as model misspecification or measurement error, among others. The comparison of the presented models using the Information criteria shown in Tables 5.3 and 5.4 revealed that different criteria may select different models, but the differences are immaterial. Specifically, the Holling type II model outperforms the Crowley Martin model in terms of AIC and Bayesian information criterion (BIC), whereas the Crowley Martin model outperforms in terms of Deviance information criterion (DIC), Watanabe information criterion (WAIC) and leave one out cross validation information criterion (LooIC). The information criteria values, on the other hand, do not deviate significantly between the deterministic and stochastic approaches. The information criteria of the stochastic models, on the other hand, have over 100 units lower values than those of the deterministic models, as shown in Tables 5.3 and 5.4. This finding reflects the fact that stochasticity absorbs noise and allows for more accurate parameter estimation.

5.3.2.3 Computational method comparison

The computation methods used in this work are based on different principles but they appear to provide robust estimates in this type of data Tables 5.5, 5.6, 5.7 and 5.8. On the one hand, there is the HMC method, which allows for the targeted exploration of the sample space by using the system's Hamiltonian to provide samples from the posterior distributions of the parameters of interest. The ADVI meanfield and fullrank methods, on the other hand, provide samples from the parameter posteriors after minimizing the Kullback–Leibler divergence between the target distribution and a family of candidates parametric distributions. The HMC method is more time consuming, but it provides robust estimates based on its convergence properties, whereas the ADVI meanfield and fullrank methods provide close results while being time efficient. As depicted in Fig. 5.2 and 5.3 VBI, both methods provide efficient posterior predictive curves that succeeded in adapting the variation of the data, particularly in the stochastic models case.

Table 5.4 Model selection criteria for the four models applied in the stochastic approach.

Stochastic case				
2 Predators				
	Holling II	Beddington	Crowely	Hassell
AIC	**210.12**	213.22	212.91	213.13
BIC	**219.68**	224.69	224.39	224.6
DIC	232.71	232.87	**232.31**	232.64
LooIC	258.90	259.89	**257.71**	258.49
WAIC	234.48	234.31	**233.62**	234.11
3 Predators				
	Holling II	Beddington	Crowely	Hassell
AIC	**196.04**	198.98	198.73	198.74
BIC	**205.60**	210.45	210.2	210.21
DIC	219.23	219.45	**219.16**	219.16
LooIC	246.07	245.89	**243.69**	247.04
WAIC	217.63	217.56	**216.94**	217.09
4 Predators				
	Holling II	Beddington	Crowely	Hassell
AIC	**227.79**	230.39	230.19	230.29
BIC	**237.35**	241.87	241.66	241.76
DIC	254.82	254.51	**254.21**	254.36
LooIC	282.13	285.71	**279.1**	282.59
WAIC	250.06	250.74	**249.82**	250.31

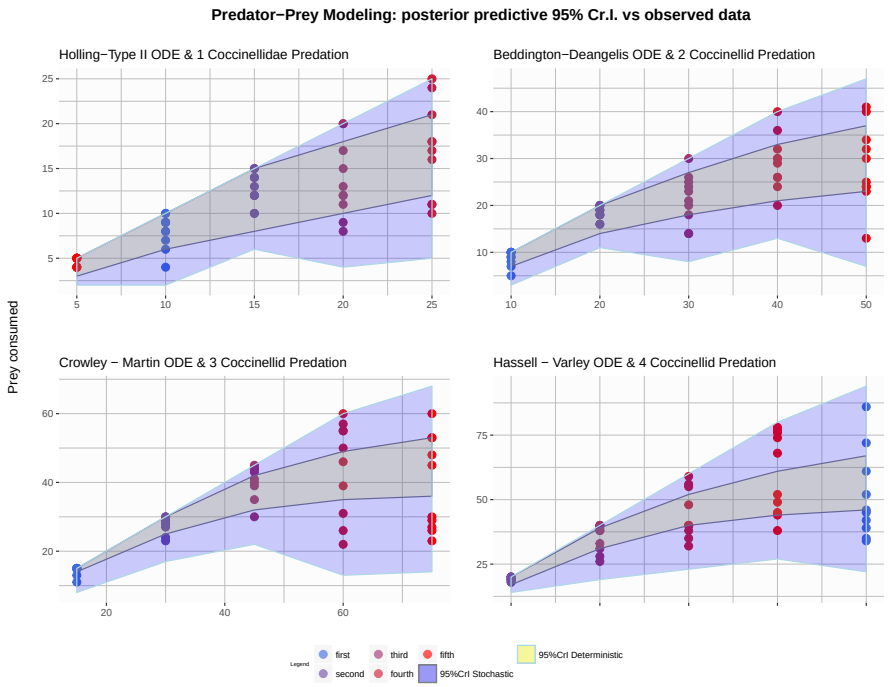

Figure 5.2 95% Cr.I. of the posterior predictive values of a. Stochastic Binomial model that contains OU process (blue colored) b. Deterministic Binomial Model (yellowish colored) vs the observed number of consumed prey (points). The figures are based upon the outcome of the HMC sampler which has better convergence properties.

Table 5.5 Posterior summaries for the Holling II ode model using the stochastic approach. In each row we report the HMC, the ADVI-meanfield and ADVI-fullrank estimates respectively.

Stochastic model for Holling type II: 1 predator							
HMC		a	T	dev†	ϕ	σ^2	ψ
	mean	0.67	0.16	159.76	2.49	5.26	1.01
	sd	0.16	0.06	10.03	2.27	4.66	0.19
	2.5%	0.39	0.04	141.8	0.46	1.22	0.68
	97.5%	0.98	0.29	181.27	8.79	18.24	1.43
	n_eff‡	3740	3208	10606	4457	4460	3085
	Rhat	1	1	1	1	1	1
ADVI meanfield ELBO: -134.77	-134.77	a	T	dev†	ϕ	σ^2	ψ
	mean	0.57	0.2	158.15	2.71	5.72	1.06
	sd	0.04	0.02	9.3	0.18	0.38	0.1
	2.50%	0.49	0.17	142.14	2.37	5.02	0.88
	97.50%	0.65	0.25	177.46	3.07	6.54	1.26
ADVI fullrank ELBO: -134.02	-134,02	a	T	dev†	ϕ	σ^2	ψ
	mean	0,42	0,11	159,87	4,98	11,01	1,11
	sd	0,06	0,05	10,32	2,57	5,52	0,1
	2,50%	0,3	0,05	141,91	1,79	3,97	0,94
	97,50%	0,54	0,23	181,87	11,3	25,58	1,31

† deviance of the model given the data.
‡ effective sample size.

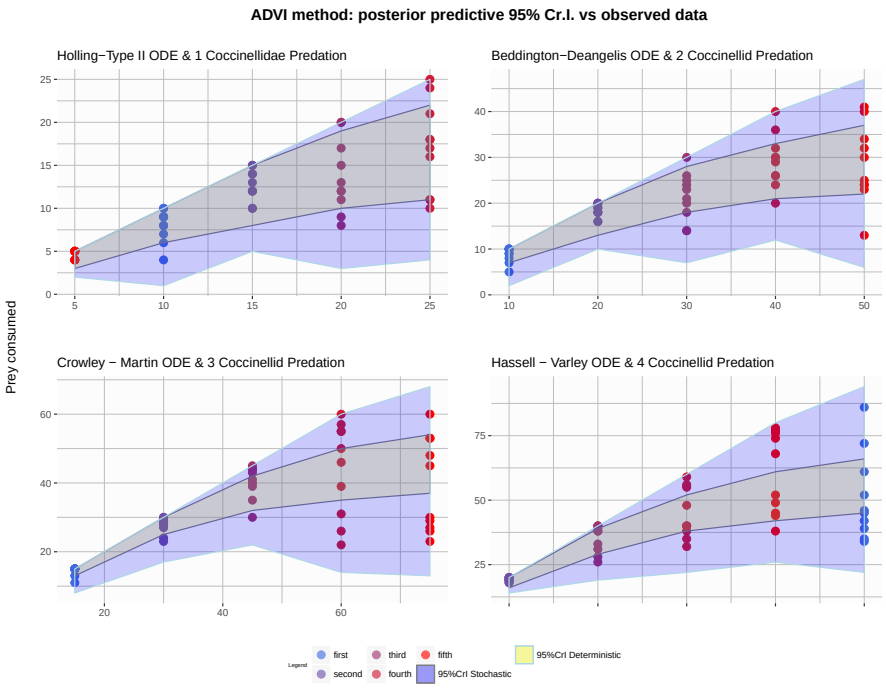

Figure 5.3 95% Cr.I. of the posterior predictive values of a. Stochastic Binomial model that contains OU process (blue colored) b. Deterministic Binomial model (yellowish colored) vs the observed number of consumed prey (points). The figures are based upon the outcome of the ADVI fullrank method which is both fast and gives reliable results.

Table 5.6 Posterior summaries for the Beddington DeAngelis ode model using the stochastic approach. In each row we report the HMC, the ADVI-meanfield and ADVI-fullrank estimates respectively.

Stochastic Beddington-DeAngelis model: 2 predators								
HMC		a	T	c	dev†	ϕ	σ^2	ψ
	mean	0.31	0.23	0.16	201.22	4.5	7.94	0.88
	sd	0.03	0.04	0.14	10.8	2.83	5.03	0.15
	2.5%	0.25	0.13	0.01	181.83	1.45	2.49	0.63
	97.5%	0.37	0.31	0.53	223.93	11.94	21.27	1.21
	n_eff‡	41114	17406	28785	13828	17190	15545	4651
	Rhat	1	1	1	1	1	1	1
ADVI meanfield ELBO: -163.152		a	T	c	dev†	ϕ	σ^2	ψ
	mean	0.3	0.23	0.11	198.67	3.27	5.98	0.92
	sd	0.02	0.02	0.08	9.65	0.19	0.39	0.08
	2.50%	0.27	0.2	0.03	181.29	2.91	5.26	0.77
	97.50%	0.34	0.27	0.31	218.9	3.64	6.75	1.09
ADVI fullrank ELBO: -162.07		a	T	c	dev†	ϕ	σ^2	ψ
	mean	0.31	0.19	0.27	205.95	2.53	4.28	0.84
	sd	0.03	0.04	0.16	11.4	0.81	1.32	0.07
	2.50%	0.26	0.12	0.05	186.19	1.33	2.33	0.71
	97.50%	0.37	0.29	0.66	230.74	4.4	7.45	0.97

† deviance of the model given the data.
‡ effective sample size.

Table 5.7 Posterior summaries for the Crowely Martin ode model using the stochastic approach. In each row we report the HMC, the ADVI-meanfield and ADVI-fullrank estimates respectively.

Stochastic Crowely-Martin model: 3 predators							
HMC		a	T	c	dev†	ϕ	ψ
	mean	0.34	0.21	0.08	186.73	3.54	1.23
	sd	0.03	0.04	0.07	9.87	2.27	0.18
	2.5%	0.29	0.13	0	169.39	1.16	0.91
	97.5%	0.4	0.3	0.26	207.82	9.64	1.62
	n_eff‡	40791	21121	22739	14686	16950	5215
	Rhat	1	1	1	1	1	1
ADVI meanfield ELBO: -168.926		a	T	c	dev†	ϕ	ψ
	mean	0.33	0.23	0.05	185.37	2.64	1.23
	sd	0.02	0.02	0.02	9.88	0.15	0.11
	2.50%	0.29	0.2	0.02	168.66	2.35	1.01
	97.50%	0.36	0.27	0.11	206.41	2.95	1.45
ADVI fullrank ELBO: -168.53		a	T	c	dev†	ϕ	ψ
	mean	0.35	0.17	0.16	196.89	2.22	1.04
	sd	0.03	0.04	0.08	11.69	0.7	0.08
	2.50%	0.29	0.1	0.04	176.41	1.14	0.89
	97.50%	0.4	0.25	0.35	221.09	3.94	1.2

† deviance of the model given the data.
‡ effective sample size.

Table 5.8 Posterior summaries for the Hassell Varley ode model using the stochastic approach. In each row we report the HMC, the ADVI-meanfield and ADVI-fullrank estimates respectively.

Stochastic Hassell-Varley model: 4 predators								
HMC		a	T	m	dev†	ϕ	σ^2	ψ
	mean	0.2	0.07	0.92	218.29	4.25	9.69	1.14
	sd	0.01	0.01	0.07	10.03	2.49	5.66	0.15
	2.5%	0.17	0.04	0.75	200.58	1.46	3.3	0.87
	97.5%	0.23	0.09	1	239.71	10.81	24.55	1.48
	n_eff‡	55277	18793	36423	14076	18540	17297	7407
	Rhat	1	1	1	1	1	1	1
ADVI meanfield ELBO: -193.147	-193.147	a	T	m	dev†	ϕ	σ^2	ψ
	mean	0.19	0.07	0.94	217.75	3.19	7.08	1.11
	sd	0.01	0.01	0.05	10.25	0.19	0.46	0.1
	2.50%	0.17	0.06	0.83	200.4	2.84	6.25	0.95
	97.50%	0.21	0.08	0.99	239.23	3.56	8.01	1.32
ADVI fullrank ELBO: -192.88	-192.88	a	T	m	dev†	ϕ	σ^2	ψ
	mean	0.2	0.06	0.86	223.53	2.95	6.7	1.13
	sd	0.01	0.01	0.08	11.97	1.03	2.28	0.09
	2.50%	0.17	0.04	0.65	203.16	1.47	3.39	0.97
	97.50%	0.22	0.08	0.96	250.01	5.55	11.88	1.31

† deviance of the model given the data.
‡ effective sample size.

5.3.3 Chapter summary

In terms of predator-prey systems, we have several issues to deal with, as well as some concerns for future research. To begin, modelling the consuming rate of prey in such systems aids in understanding and describing the rate of change in the prey population while accounting for inter-individual interferences not only between species but also within the same species. The presence of noise due to the nature of ecological data, as well as the potential failure of models to detect any kind of systematic source of variation in such data, is an additional challenge in modelling.

5.3.4 Computational methods overview

To apply the modelling class we suggested in the predator prey dynamic system features, we use the Bayesian paradigm as well as some contemporary computational approaches such as the HMC, ADVI-meanfield, and ADVI-fullrank. In such data schemes, ADVI techniques provide robust estimates of all the parameters of the models we study that are close to the HMC method, gaining in time efficiency. Because of its convergence properties, the HMC method produces robust estimates in all model cases. The STAN software platform allows for a parallel approach to finding solutions to the ordinary differential equations required during iterative sampling procedures in all methods.

5.3.5 Distribution of the data

Furthermore, we recommend the Binomial distribution for data generation because it is appropriate for observed counts of prey consumed during the study time intervals. The Binomial distribution probability is linked to and derived from the instantaneous prey abundance. The most common link is the instantaneous ratio of consumed prey density, which leads to the case of deterministic models.In this study, we proposed a link to a OU process that is centered and reduced to (in the absence of diffusion) the instantaneous ratio of consumed prey density. Thus, stochastic models are a generalization of deterministic models in this context, accounting for data variation due to noise sources.

5.3.6 Model comparison

Furthermore, we use the Information criteria to compare models not only between different ode methods, but also between deterministic and stochastic models. Using the leave one out cross validation technique, we can assess both the models' goodness of fit to data and their ability to make predictions on new data. Finally, we plot the predictive posteriors alongside the observed data points to see how well the suggested models predict.

5.4 Future research

Future research should focus on creating full predictive model selection using proper scoring rules. Another area of interest is investigating and approximating optimal control techniques for solving predator-prey system constraints using the Bayesian paradigm.

Chapter 6

Estimating and mapping transmision risk for vector-bone diseases

6.1 Introduction

VBDs are human illnesses caused by parasites, viruses, and bacteria spread by vectors (W.H.O. 2021). Mosquitoes, ticks, sandflies, triatomine bugs, tsetse flies, fleas, black flies, aquatic snails, and lice have all been identified as disease vectors (Müller et al. 2019). The emergence and spread of VBDs in Europe are a result of disease-causing biotic, abiotic, and socioeconomic factors (Semenza 2015). In addition, globalization and environmental change, socioeconomic and demographic factors, and health system capacity are all major contributors to VBDs, which can also operate as epidemic progenitors. Thus, tracking changes in these factors can aid in anticipating, or even forecasting, an outbreak of infectious diseases (Semenza 2015). Mosquitoes (Diptera: Culicidae) are the most important hematophagous arthropod vectors of numerous human illnesses. Malaria, West Nile virus, yellow fever, chikungunya, Zika virus, and Japanese encephalitis are examples of mosquito-borne illnesses (MBDs) (Schaffner et al. 2013, Calzolari 2016). Given the invasive success of numerous mosquito species as a result of global human mobility and trade (Kerkow et al. 2020), the development of early warning and control strategies appears critical (Pergantas et al. 2017).

Malaria is a parasitic infectious disease (VBD) spread by the bite of infected anopheleline mosquitos. It is responsible for an estimated 219 million illnesses and over 400,000 deaths globally each year. The vast majority of deaths occur among children under the age of five (W.H.O. 2021). Public health decision-making generally needs early warning output from systems which are based on uncertain data (Degallier et al. 2010, Kuhn et al. 2005). Considering the importance of VBDs in human health, it is imperative to work towards creating a suitable framework for an EWS which would improve our understanding of the connectivity between existing and potential vector-borne risk areas.

Given the importance of VBDs in human health, we proposed in (Pergantas et al. 2021) an early-warning system for VBDs. The current study aims to highlight tools for depicting the degree of risk of transmission of VBDs like malaria on a map of a larger understudy region like Central Greece. This visualization, together with the feedback of transmission risk measures from the epidemiological model proposed in (Pergantas

et al. 2017), contributes to the development of an integrated process that can serve as a predictive tool for public health preparedness and response. The latter is achieved by providing insights into the relationship between actual and potential risk areas as transmitted by operators.

This chapter describes the structure of the early warning system while staying within the mapping tools, as well as its application to real-world data from Central Greece. To that end, we first present the proposed model class components before describing how to generate spatial graphs depicting the derived estimates in heat map formulation using the R platform (Team 2021). The model class is a customized version of the well-known Ross-Macdonald mathematical model defined in equations (6.1) and (6.2), which estimates the average number of VBD infections in the study area using entomological, social, environmental, and geographical data. The model is made up of three components-measures that account for different but complementary severity scales. The first component is the expected number of infections caused by one infected human-host (basic reproduction rate), which accounts for the disease potential in the study area (mosquito-driven). The second component is the probability of infection, which accounts for the disease potential in the host population, which includes immigrants from malaria-endemic areas. The expected number of infected cases in the study area is the third component. Following that, we describe the R-tools that were used to generate several spatial maps that not only contain potential infection points but also depict the intensity of the aforementioned estimates of suitable risk indicators. In terms of public health, the objective of this study would be to employ the model class, as well as the associated geographical mapping of the generated estimates from this model, as an early warning system with meteorological inputs, allowing for better decision making and (VBD) prevention.

6.2 Epidemic model class for VBD

In this section, we present the previously studied host-vector model in (Pergantas et al. 2017), which integrates entomological, sociological, environmental, and geographical variables to produce estimates of three complementary risk measures of malaria resurgence within the study region of Central Greece. The basic reproduction rate of the disease (R_0), the probability of (human) infection (τ), and the number of expected (human) infections in a region ($E\left(\text{infections}\right)$) are among the risk measures. The standard entomological parameters, which are accounted for in the measure R_0, are estimated using the well-established Ross-Macdonald mathematical model (Macdonald 1952). Furthermore, the model considers the possible host population in the study region throughout the estimation of the probability τ, which is associated with immigrants from malaria-endemic locations (Smith et al. 2004).

6.2.1 Parameters of the model

The disease's basic reproduction rate, R_0 , and the probability of an individual becoming infected, $\tau = \Pr(\text{infection})$, are two of the basic model's risk measures. The R_0 in the framework of VBDs is the number of secondary disease infections caused by a single infected individual in a population that is otherwise unaffected. It is a natural threshold parameter that is ideal for disease control because an epidemic can occur only when R_0 is greater than one. This demonstrates the potential for disease spread if a single infective individual starts an outbreak. Furthermore, the probability τ is influenced by two factors. The first is concerned with potential population infection as a result of mosquito abundance, whereas the second is concerned with (human) host infections. To quantify these contributors, the reproduction rate R_0 is used for the former, while the proportion of initially infected (human) hosts, denoted by μ_0, is used for the latter (Pergantas et al. 2017). The corresponding model-based point (typically median) estimates of these measures are denoted by $\hat{R}_0$ and $\hat{\tau}$.

Furthermore, the expected number of infections in a region, say $E(\text{infections})$, is used as a third risk measure, taking into account the probability of infection weighted with the abundance of susceptible population in the study area. The probability τ and the expectation of infections $E(\text{infections})$ are two risk measures that are only meaningful in the event of introducing infected individuals into the area, as is $R_0 > 1$. The estimators' definitions of the aforementioned risk measures are presented below.

6.2.1.1 Reproduction rate of the disease R_0

According to (Pergantas et al. 2017) the formula (6.1) is used to calculate local R_{0i} estimates for the ith sample-collection station:

$$\hat{R}_{0i} = \frac{Vec_i \cdot b_i \cdot c}{r_i} \tag{6.1}$$

where Vec_i is the vectorial capacity in station i, or the expected number of infective mosquito bites that would result from all mosquitoes biting a single fully infectious person on a single day (Smith et al. 2004), and is given by:

$$Vec_i = \frac{m_i \cdot \alpha_i^2 \cdot \exp(-g_i \cdot v_i)}{g_i} \tag{6.2}$$

In (6.1) and (6.2), m_i denotes the numbers of mosquitoes in each station i; α_i the biting rate, i.e. the percentage of mosquitoes that feed on humans each day; b_i the probability a bite produces infection to a human; r_i the average daily recovery rate per day; v_i the mosquito latent period, i.e. the number of days from infection to infectiousness; g_i the mosquito mortality rate per day. Finally, with c we denote the probability a bite

turns a susceptible mosquito to infected, which for our analysis is set to the constant value of 0.5. The parameters α_i, b_i, r_i and v_i were sampled from suitable distributions according to the relevant literature (Smith and McKenzie 2004), whereas g_i changes with the temperature levels (Smith et al. 2004), which are currently represented by monthly means of temperature.

6.2.1.2 Host infections due to migration

In addition to the entomological part of the proposed model, estimation of the external host component due to the migration is embedded into the risk parameter calculations, by utilizing an exponential kernel function, W_{ik}, of the form:

$$W_{ik} = \alpha_0 \cdot \exp(-\alpha_0 \cdot d_{ik}) \tag{6.3}$$

W_{ik} is used to model the spatial part of the potential inflected hosts. In (6.3) d_{ik} denotes the distances from larvae areas, measured during the three periods of potential hosts' monitoring (k=1,2,3). Subsequently, the estimation of the external host component due to the migration is approximated by: $\hat{\mu}_{0i} = \sum_{k=1}^{3} \mu_{0ik} \cdot W_{ik}$. This estimated proportion of initially infected hosts, $\hat{\mu}_{0i}$, is then multiplied by a predetermined incidence rate derived from a sensitivity analysis (Pergantas et al. 2017).

6.2.1.3 Probability of getting infected

Finally, according to (Pergantas et al. 2017), estimating probability of becoming infected $\hat{\tau}$, in the event of a local outbreak is achieved by solving the non-linear equation:

$$1 + \hat{\mu}_{0i} - \hat{\tau}_i - \exp\left(-\hat{\tau}_i \cdot \hat{R}_{0i}\right) = 0 \tag{6.4}$$

which only applies for $\hat{R}_{0i} \geq 1$. For completeness shake, we set $\hat{\tau}_i = 0$ when $\hat{R}_{0i} < 1$.

6.2.1.4 Number of expected infections

The number of expected infections over unit of time is given by:

$$E\,(\text{infections}) = \Pr\,(\text{infection}) \times (\#\text{ of susceptibles}) \tag{6.5}$$

Our collaborative work on developing an augmented semi-automatic early warning system tool has been published in (Pergantas et al. 2021). The estimating and mapping graphs were created in stages using the free and open-source R software environment (Team 2021). These steps are described in detail in the section 6.3.

6.3 Methodology for creating risk measurement maps

To perform the calculations and draw the figures, a variety of R library packages are used. The 'dplyr' package is used specifically to facilitate data frame manipulation, and *ggmap* (Kahle and Wickham 2013) provides the main methods for generating the data graphs in this work. The latter package, when combined with the ability to retrieve a large collection of static maps from various online sources (e.g., Google Maps and Stamen Maps) via package embedded procedures, produces an effective visualization of the parameters of interest. We principally used the Stamen maps option in this work. Initially, the *get_stamenmap* function is used to create a bounding box (bbox) object whose boundary is determined by the geographical coordinates of the study region in order to visualize any map graph. The dimensions of the output map are determined by a vector that consists of the minimum and maximum latitude changed (subtracted and added) by 10% of its total range for the bottom and top boundries, and the minimum and maximum longitude changed (subtracted and added) by 10% of its total range for the left and right boundaries, respectively. The zoom level on the region and the 'maptype' between terrain, terrain-background, terrain-labels, terrain-lines, toner, toner-2010, toner-2011, toner-background, toner-hybrid, toner-labels, toner-lines, toner-lite, or watercolor are also options for the bbox object. The terrain option is the default and is used extensively in this work. Another option is toner-background, which converts the terrain background to white and the sea area to black. The map graphs presented in this work include risk measurement estimates and can be divided into two main categories based on graph form, as described in the following two paragraphs.

6.3.1 Measurement based point maps

The *qmplot* function, which creates a quick overview of maps and data points, is one of the tools used in this first category. The data points in this section are spatial points that reflect measurements on terrain based maps in (Figs. 6.1-6.3). The longitude and latitude values of the spatial points of interest, the data frame containing the data, the zoom on the map, the map type as described in the previous paragraph, and geometry of the output "geom" which is by default set to "point" option are some essential arguments of the *qmplot* function. Furthermore, the *ggmap* function can be used in conjunction with the *geom_point* function to add points to a region map. The former function is required to create the previously mentioned "bbox" on map-type determined manually by the user. To insert points at specific geographical coordinates, the latter function is necessary. The arguments used in *geom_point* are the "data" which define the data frame containing coordinates and measures data and the "mapping" which defines the aesthetics of the points that comprise of the "x, y" axis values, the "col" and "size" which

define which are the parameters to change the color and size of the points based on their value intensity. In our example, toner-background based map type are used in (Figs. 6.5-6.7), and the size and color vary depending on the values of (R_0), (τ), and (E). Other useful 'ggmap' library commands worth mentioning are "scale_color_gradient" and "scale_color_distiller," which provide a color gradient defined either by the color names or by a palette and its direction, associated with the range of parameter estimates we use each time.

6.3.2 Kernel density based maps

The second category contains the *stat_density2d* function, which estimates kernel density in two dimensions using an axis-aligned bivariate normal kernel function (Figs. 6.8-6.10). By quantifying the contribution of each data point on a map region, this density function generates a continuous surface. This contribution is smoothed out from a single point into the space around the point. On a grid, the kernel density estimate is given by:

$$\hat{f}(x,y) = \frac{\sum_i \phi\left(\frac{x-x_i}{\sigma_x}\right) \cdot \phi\left(\frac{y-y_i}{\sigma_y}\right)}{n \cdot \sigma_x \cdot \sigma_y}, \tag{6.6}$$

where density ϕ denotes the standard normal distribution and $\mathrm{diag}\left({\sigma_x}^2, {\sigma_y}^2\right)$ denotes the bandwidth diagonal matrix, which controls the amount and direction of smoothing induced. The bandwidth serves as the covariance matrix of the bivariate normal kernel (Venables and Ripley 2013).

By accumulating the intersections of the various regions, the output surface indicates where point characteristics are concentrated. The method of density computation is determined by the bandwidth, which has a default search radius. Furthermore, the "bins" option is a control parameter that sets the number of contour levels. We used 100 "bins" for the construction of Figs. 6.8-6.10.

The *ggmap* function which plots the raster object created by the aforementioned *get_stamenmap* function is required for the *stat_density2d* function to run. The later function requires the data frame ("data" option) containing the characteristic values, the aesthetics ("aes" option), which should include the (x, y) geographical coordinates, and the fill parameter, which specifies the range of the color gradient (here "nlevel" is determined by "n" the number of data) and the "bins" argument. In addition, the geometrical option "geom" which here is suitable for polygon shapes to form contour lines. The Supplement C contains code for implementing all of the methods described in the paper.

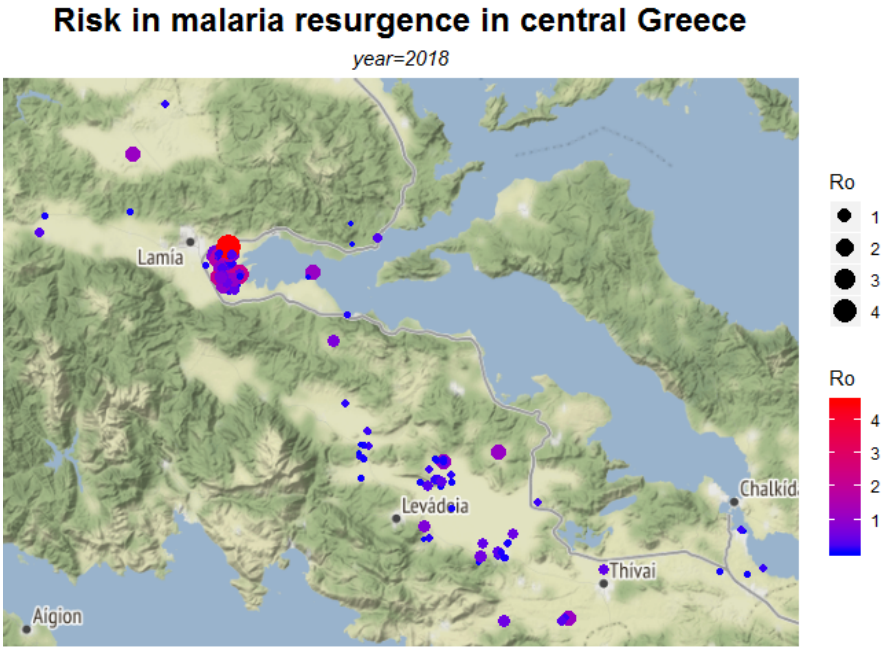

Figure 6.1 Map depicting the risk of malaria transmission in the form of R_0.

6.4 Malaria transmission in Greece: a real-life application

The research was conducted in eight municipalities of the Prefecture of Central Greece (Pergantas et al. 2017). The research area was approximately 406.000 hectares (ha). For mosquito data collection, CO_2 traps were used at each surveillance site.

Furthermore, female adult mosquitoes, which are malaria transmission vectors, were included in the study. The traps were placed in ten regions, totaling 393 sample-collection points on the fringes of cities and villages in the Prefecture of Central Greece. Canals, rice paddocks, and tanks were among the breeding places, with the majority of them previously verified for the presence of mosquitoes using drones. Between 6/3/2018 and 29/8/2018, the traps were set and checked every 10 to 15 days.

The morphological evaluation of mosquito samples was used to identify the genus (Pergantas et al. 2017). The species found in the study region, in particular, account for roughly 90% of all Anopheles species. As a result, all positive Anopheles mosquito larvae samples are thought to be malaria vector larvae. Temperature data were collected from the National Observatory of Athens (NOA) in 10 meteorological stations.

The NOA operated ten meteorological stations in the study area that recorded temperature data. The average temperature values are used in the EWS model with a 30-day interval, beginning on the first Saturday of each month between March and August of 2018. The inverse distance weighting (IDW) interpolation method (Lukaszyk 2004) was additionally used to estimate the average temperatures at the locations where no

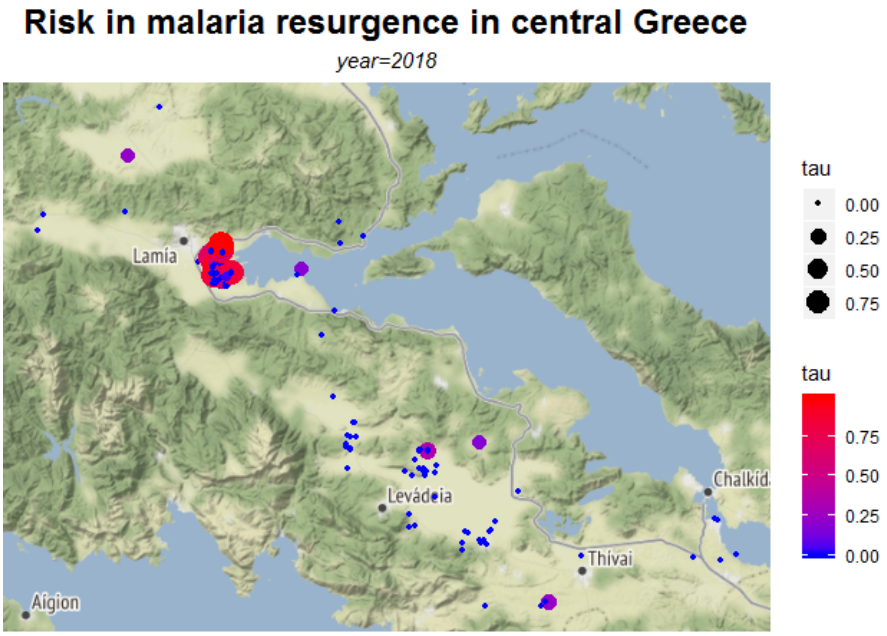

Figure 6.2 Map of risk of malaria transmission, computed by the EWS model (τ estimates).

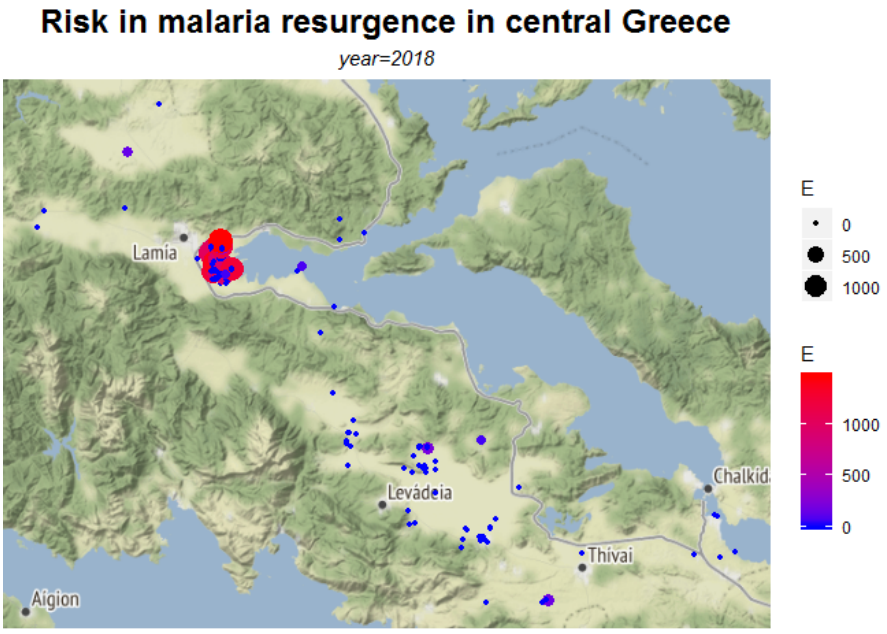

Figure 6.3 Map of risk of malaria transmission, computed by the EWS model (E estimates).

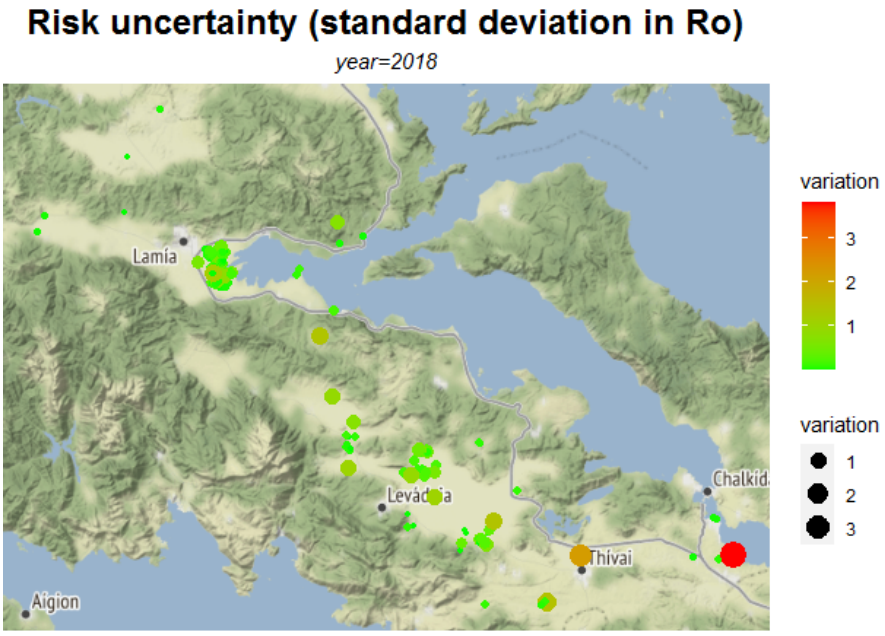

Figure 6.4 Uncertainty in R_0, in the form of the standard deviation of R_0 (based on 1000 samples).

measurements were available.

6.4.1 Estimation of the risk measures

Figures 6.1-6.3 provide risk maps for the three risk parameters of interest. The intensity of each risk measure varies with color on these maps, which are based on the geographical coordinates of the data points.

In Fig. 6.1 the risk of malaria transmission expressed by the median basic reproduction rate, $\hat{R_0}$, for each area is depicted on study-region's map. Besides, Fig.6.2 depicts the estimated probability of infection, τ. The number of predicted infections E, as calculated by the geographical prediction model, is shown in Fig. 6.3.

These visualization maps were made with the *qmplot* function from the 'ggmap' library package. The *ggmap* package also includes alternative presentations based on other functions, such as the *ggmap* along with the *geom_point* or the *stat_density2d* functions respectively. Figs. 6.5 to 6.10 includes alternative visualizations of our results, including maps of point estimates using alternative terrain representations or heatmaps respectively and different options may suit different users based upon their needs.

Values of the R_0 above (Fig. 6.1), indicate where the greatest potential for risk

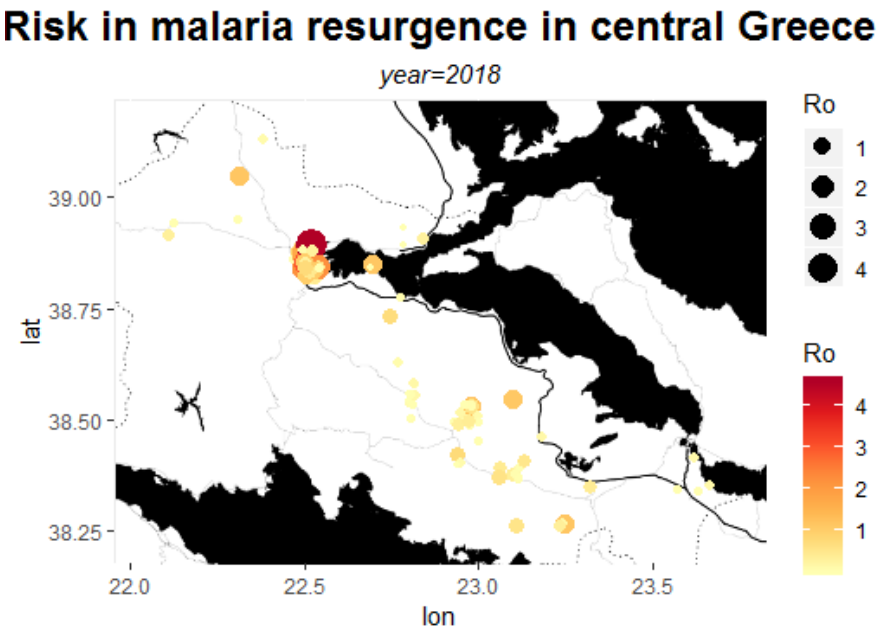

Figure 6.5 Point estimate map of risk of malaria transmission (R_0 estimates).

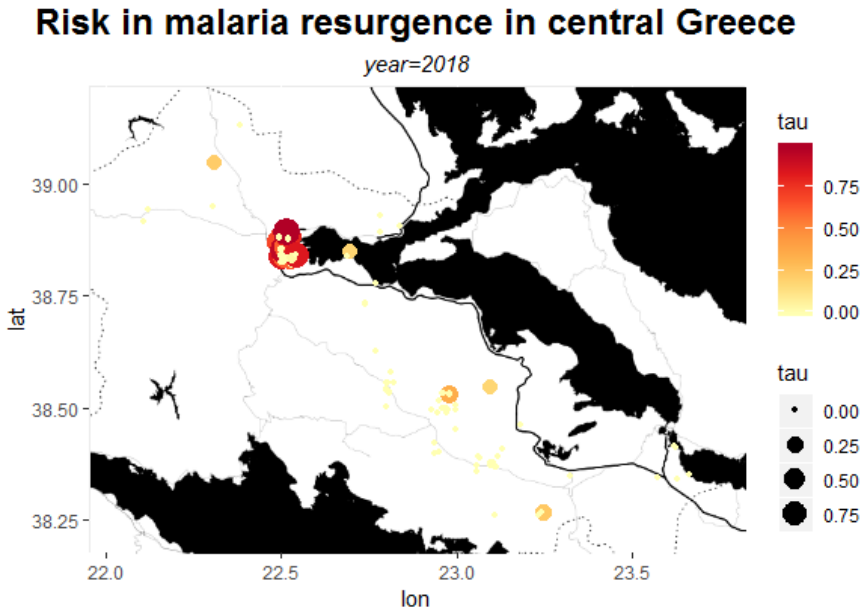

Figure 6.6 Point estimate map of risk of malaria transmission (τ estimates).

is located. The highest risk is primarily located in the area of Lamia, with R_0 reaching values as high as 4. The probability of getting infected from low to high (zero to 0.75) was depicted in the map of Fig. 6.2 and the Lamia region was found to have the highest risk. However, the map also suggests non-negligible probability of infection (τ estimates between 0.25 and 0.5) in the largely dispersed rural areas of the Prefecture.

As these rural areas have low populations, the expected number of infections are relatively low as shown Fig. 6.3, revealing the complementary characteristics of the different risk measures. The highest number of potential infections is concentrated in the wider region of Lamia.

In addition to risk maps based on estimated parameters such as the basic reproduction rate R_0, our approach allows for the presentation of the associated uncertainty of the estimated parameters (or functions thereof) by depicting the associated variability, such as the variance or standard deviation. For example, the risk map of malaria transmission based on R_0 (Fig. 6.1) can be naturally integrated with a parameter variability map (see Fig. 6.4) to give a more robust tool for monitoring VBDs transmission. This type of combined reporting can be potentially applied to the other measures of risk assessment. Perhaps more importantly, it reveals knowledge gaps since high uncertainty suggests that further sampling is required in those areas in order to reduce this variability. In order to examine the potential for malaria resurgence in the study areas, our

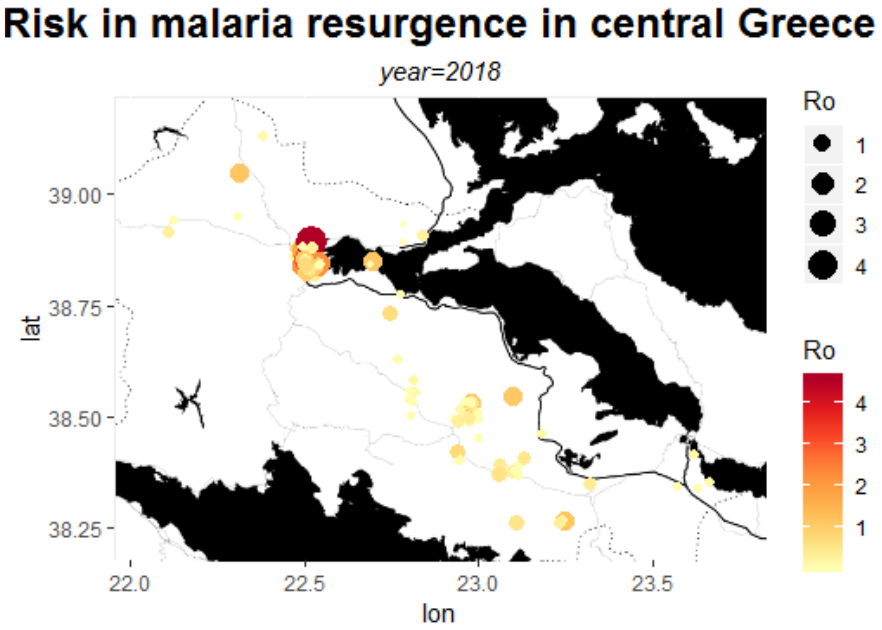

Figure 6.7 Point estimate map of risk of malaria transmission (E estimates).

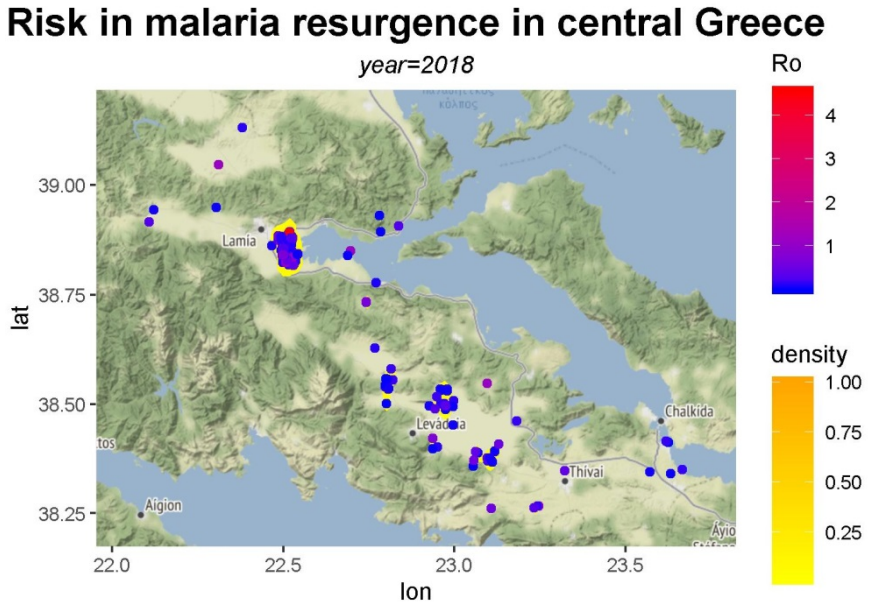

Figure 6.8 Heat-map of risk of malaria transmission (R_0 estimates).

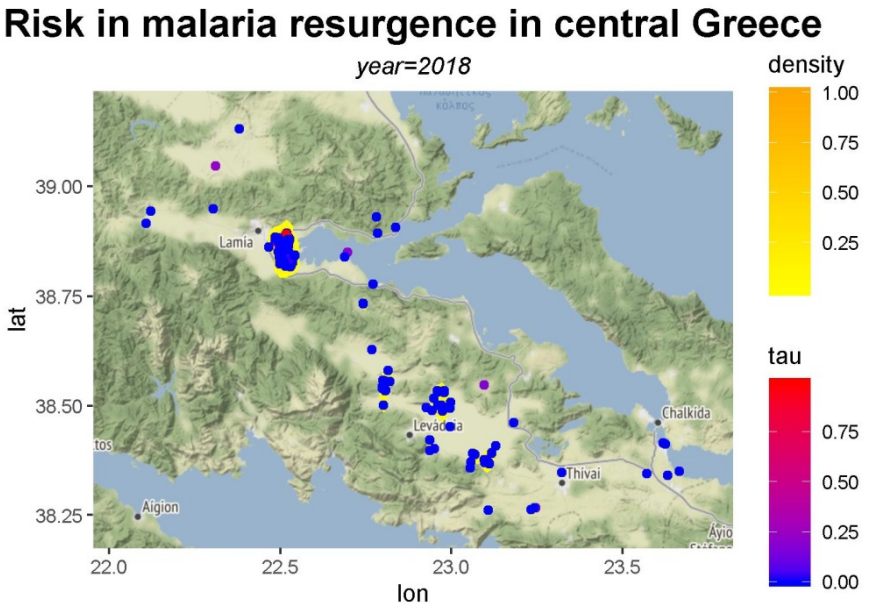

Figure 6.9 Heat-map of risk of malaria transmission (τ estimates).

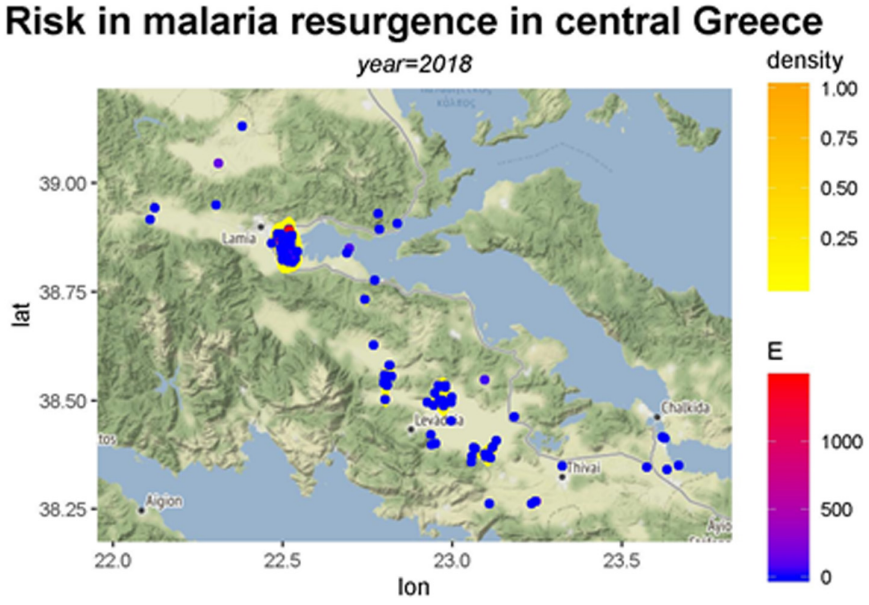

Figure 6.10 Heat-map of risk of malaria transmission (E estimates).

study suggests appropriate tools in the R platform for generating appropriate maps that can adequately describe their spatial distribution of risk. This distribution is derived using appropriate mathematical modelling in a EWS model-based framework. The corresponding EWS model, combined with the visualization of risk estimates in R-created maps, results in a semi-automatic open source tool for early warning of mosquito-borne diseases such as malaria and west Nile virus, among others.

The results in our example imply that by inspecting the generated graphs, five to six separate geographical areas with the potential for malaria resurgence can be recognized. The areas most at risk include those near Lamia, as expected due to the local rice fields, and, to a lesser extent, the lowlands of Levadeia and Thivai, where an outdated irrigation system is responsible for a huge number of Anopheles mosquito breeding sites. In general, the coexistence of people from malaria-endemic countries and those engaged in agricultural activity in places with Anopheles mosquito breeding habitats, such as paddies (e.g., the city of Lamia) and irrigation canals, reduces the probability of malaria resurgence. These places have the potential to be hot-spots for the recurrence of malaria.

Besides, refugee communities have recently been hosted across Europe, including Greece (Eurostat 2020), significantly raising the danger of vector-borne disease transmission. Furthermore, (Fotakis et al. 2020) discovered various disease vectors in Greek refugee camps, indicating a possible risk factor for disease transmission. To that end using the early warning system and inspect the outcome maps for vector-borne diseases, such as the one described in this research, may be useful in preventing disease spread.

Our approach could also be utilized as a tool for efficient mosquito species control, showing time control periods, restricting mosquito expansion, and hence the possibility for disease transmission. Understanding mosquito spatial distribution is critical for public health and serves as a foundation for studies aimed at understanding their spread (e.g. (Minakawa et al. 1999, Kerkow et al. 2020, Yamasita et al. 2018)). Studies dealing with species distribution models may overlook essential aspects that drive these models, such as training data quality and critical abiotic factors (Kerkow et al. 2020). Our mosquito sampling approach included CO_2 traps. As a result, our data included only female *Anopheles* individuals, which serve as malaria vectors, excluding the possibility of sex-biased data from larvae sampling.

In turn, EWS model suggested accommodates for temperature changes, which are very essential in insect performance ((Huey and Berrigan 2001*b*, Huffaker et al. 1971)). Temperature is the most important abiotic factor influencing insect dispersal, influencing crucial components of their life cycle such as development, survival, reproduction, and life span (e.g. (Kontodimas et al. 2004, Jalali et al. 2010, Papanikolaou et al. 2013)). This has an additional impact on insect fitness, influencing population dynamics (e.g., (Kontodimas et al. 2007, Papanikolaou et al. 2014, 2019)). As a result, by accounting

for temperature variations, our model is geared toward accurate assessment of mosquito population growth.

Our method has some limitations. It is based on many types of comprehensive evidence. This type of data may or may not be easily available, and some form of estimate is frequently used. However, such issues are reflected in the uncertainty of the risk measurements, and the corresponding maps are a natural result of the suggested technique. In reality, these maps provide an opportunity because they indicate where extra sampling should take place to reduce uncertainty. Furthermore, the majority of such data are observational and do not constitute part of a randomized controlled experiment. This study is a typical example, and appropriate counterfactual scenarios, such as what would have happened if no vector control mechanism had been used, are prevalent in the field. These scenarios are based on established theory though and including the corresponding uncertainty facilitates scientifically honest reporting and leads to additional data collection as described above.

6.5 Chapter summary

The mapping tools provided by the 'ggmap' library and presented here are suitable for visualizing critical complementary risk measures on the map as well as monitoring the evolution and resurgence potential of the VBDs.

To that end, we anticipate that these mapping tools will propel the semi-automatic EWS developed in (Pergantas et al. 2021) to potentially aid in the monitoring and control of VBDs in a variety of settings.

In addition, the augmented EWS model described here can be used for optimizing the cost-effectiveness of distinct control measures and the integration of open geospatial and climatological data.

Last but not least this kind of EWS model can enhance our ability to predict the risk of disease outbreaks while climatic conditions change. The developed methods can be adapted for usage in countries with similar vector-borne disease potential. The R code producing and depicting the risk indicators is provided in the supplement C.

Chapter 7

Discussion

In the order shown in the previous chapters of this thesis, this Chapter summarizes the most important findings and conclusions of this work. It also includes suggestions for future research and the expansion of current work.

7.1 Summary of most important findings and conclusions

7.1.1 Statistics in Demography

7.1.1.1 Deterministic demography

The addition of uncertainty to the demographic cohort measures via bootstrap allows for hypothesis testing. To that end, studying demographic statistics in real data shows that temperature has a significant impact on *T. granarium* population growth. Specifically, at $40°C$ the value of the intrinsic rate of increase is negative, indicating that at this temperature the population tends to extinction, although *T. granarium* is considered a highly heat-tolerant species (Lindgren et al. 1955, Lindgren and Vincent 1959). At 30 and $35°C$ the positive values of the intrinsic rate of increase indicates that in this temperature range *T. granarium* is able to increase its population size, as well as its potential to spread, becoming more harmful in stored-products. Furthermore, while the mean generation time at $30°C$ is significantly lower than that at $35°C$, but not in the other demographic factors, this is biologically meaningful because the net reproduction rate is determined by cohort survival, which is lower for *T. granarium* at $35°C$. However, because the values of the intrinsic rate of increase and the doubling time did not change much, we expect that the insect's growth rate will be similar between these temperatures. The minimum and maximum temperatures for *T. granarium* population increase, according to the Briere model fit, are around $18.44°C$ and $40.00°C$, respectively. *T. granarium* can multiply in this temperature range, but the subsequent decrease at higher than optimal (at $34.52 \circ C$) temperatures is most likely due to the temperature's determinant effect on its survival and reproductive capacity. This is critical for the management of this species, given its economic importance and its global spread, as well as its mass-rearing, which allows for efficient breeding in the insectary (Carey and Vargas

1985). Furthermore, female reproductive value, or the contribution an individual of a certain age will provide to future generations, grows until a certain age. This is due to the early mortality of *T. granarium*'s pre-reproductive age classes and the subsequent increase in age-specific fecundity. Following that, a fall in age-specific fecundity has a negative influence on reproductive value, which decreases to zero as one gets older. Individuals of around 63, 42, and 21 days of age attain their optimum reproductive potential at 30, 35, and $40°C$, respectively. The expected remaining lifetime decreases until a specific age of 35 and $40°C$ due to early mortality, then increases due to decreasing mortality, and finally decreases. The expected remaining lifetime at $30°C$, on the other hand, is characterized by a continuous decrease due to no notable early mortality.

7.1.1.2 Stochastic demography

In the stochastic approach initially, we model the process of *T. granarium* laying eggs by a Zero Inflated Poisson distribution. Statistical learning for models of this kind represents a non-standard problem due to irregularities in the likelihood function and adopting a sampling-based approach to inference such as MCMC offers a substantial advantage, including the ability to estimate the complete posterior distribution of the Poisson rate and the probability of excess zeros. The separation of this probability at $40°C$ from the other two temperatures is readily evident by basic eye inspection, and this is a desired aspect of the suggested statistical study. Investigating influential individuals to this stochastic approach using the posterior density, indicating that the model appears to accommodate all of the individual data reasonably well because no significant deviation from the majority of the observations was observed. Secondly, survival analysis techniques were utilized in order to examine the duration (in a life cycle generation) both (i) until the event of "death" and (ii) until the event of *T. granarium* females become active and lay their first egg which gives rise to an independent censoring mechanism. Using the Kaplan-Meier estimators the means of the survival time until death of *T. granarium* are 62.88, 34.25 and 15.61 days while their medians diminish rapidly, fact that reflects the high mortality rate at temperatures 35 and 40. Furthermore, the means of the time until first egg release are 71.70, 46.90 and 81.30 days while the medians decrease when the temperature rises from 30 to $35°C$. The probability of *T. granarium* laying the first egg does not cross the 0.5 line at $40°C$, so the median time to first birth cannot be estimated. The final results are most likely influenced by the fact that survival time decreases significantly as temperatures rise from 35 to $40°C$. Thirdly, after several popular models comparison, the parametric approach yielded specific parametric distributions (Loglogistic and Lognormal, respectively) for the underlying biological process. These results can be used to investigate the universality of these distributions in future research by examining the parametric forms of these types of duration for

related species.

7.1.2 Modelling developmental rates

Constructing and comparing non-linear non-nested models with varying numbers of parameters and truncated mean structures, as well as excessive-zeros in data, is not an easy task when it comes to insect developmental rates.

7.1.2.1 Computational methods overview

We employ the Bayesian paradigm, as well as some recent computational approaches such as the HMC, ADVI-meanfield, and ADVI-fullrank, to address not only irregular and truncated mean structures, but also the uncertainty of zero generation in the data. Although ADVI techniques are gaining popularity in the scientific community due to their fast and computationally inexpensive approximations to the posterior distributions, they do not provide robust estimates of all the parameters of the models we study. The HMC method, on the other hand, gives robust estimates even under these specific model and data structure conditions.

7.1.2.2 Distribution of the data

Furthermore, for the data generation scheme, we suggest the Gaussian and the Inverse Gamma distributions. The Gaussian option provides sensible estimates that can be used even though the data has a lot of zeros. Inverse Gamma, on the other hand, not only naturally models developmental rates, which are characterized as the reciprocal of positive real values, but also provides variance adaptivity across temperature fluctuations. Also for each ecological model we define the Zero Inflated Inverse Gamma density so as to model data with an excessive number of zeros. When comparing models involving Gaussian and Inverse Gamma or Zero Inflated gamma distributions, we find that the second performs better in non-zero data cases, while the first performs better in all but the Analytis model case.

7.1.2.3 Model comparison

Furthermore, we address the model comparison difficulty by utilizing the Information criteria not only to assess model goodness of fit to data while accounting for model complexity, but also to assess models' capacity to generate predictions on new data using the leave one out cross validation technique. Furthermore, we use marginal likelihood estimates of the several models to assess which one is best supported by the data. Finally, we plot the projected posteriors alongside the observed data points to see how well the suggested models predict.

7.1.2.4 BMA performance

The predictive bias of the weighted models, as well as the uncertainty about the weights, affect BMA results using Information criteria as weights. As a result, the BMA approach does not provide estimates that differ from the best-performing models.

7.1.2.5 R-package for modelling developmental rate

We suggest 'DRIN', a modern R package that combines the various nonlinear ecological functions and data distributions used in previous sections into a single user-friendly function for estimating model parameters of interest and inferring for insects developmental rates. The model code utilized, which makes use of STAN software capabilities, can address not only the issues of limited parameter sampling space and reduced ecological functions used, but also the observed data excesses zeros. Finally, it provides posterior predictive and deviance estimates, which may be used to check fit and compare models.

7.1.3 Modelling predator prey systems with predator interactions

The study of predator-prey dynamics is critical in ecology for understanding how to describe and predict population performance, as well as how to manage invasive species.

7.1.3.1 Distribution of the data

We used the Binomial distribution for data generation because it is appropriate for observed counts of prey consumed during the study time intervals. The Binomial distribution probability is linked to and derived from the instantaneous prey density which is described by popular ecological models based on ode. The most common link is the instantaneous ratio of consumed prey density, which leads to the case of deterministic models. In this study, we proposed a link to a OU process that is centered and reduced to (in the absence of diffusion) the instantaneous ratio of consumed prey density. Thus, stochastic models are a generalization of deterministic models in this context, accounting for data variation due to noise sources. Furthermore, the widely used Holling type II model is used as the null ecological model, as opposed to more recent ode models that account for predator-predator interactions.

7.1.3.2 Model comparison

Moreover, we use the Information criteria to compare models not only between different ode methods, but also between deterministic and stochastic models. Using the leave one out cross validation technique, we can assess both the models' goodness of fit to data and their ability to make predictions on new data. Finally, we plot the predicted

posteriors alongside the observed data points to see how well the suggested models predict.

7.1.3.3 Computational methods overview

Furtermore, to apply the modelling class we suggested in the predator prey dynamic system features, we use the Bayesian paradigm as well as some contemporary computational approaches such as the HMC, ADVI-meanfield, and ADVI-fullrank. In such data schemes, ADVI techniques provide robust estimates of all the parameters of the models we study that are close to the HMC method, gaining in time efficiency. Because of its convergence properties, the HMC method produces robust estimates in all model cases. The STAN software platform allows for a parallel approach to approximate solutions to the ordinary differential equations required during iterative sampling procedures in all methods.

7.1.3.4 Real data results

In the case of a single predator (P=1), all models are equivalent to the Holling type II model, and the predator's mean handling time falls within the 95% Cr.I. intervals (0.15h, 0.28h) and (0.21h, 0.38h) for the deterministic and the stochastic case respectively. Moreover, there is no statistically significant deviation in any of the four ecological models tested. This finding suggests that predators' per capita feeding rate and predator density are roughly independent. Concerning attack rates there are no significant differences when the number of predators is less than four, but the presence of four predators results in a statistically significant decrease in attack rates in all models, both deterministic and stochastic. Estimates of variables c and m exclude zero and show no significant differences as the number of predators increased, implying that the magnitude of interference is not negligible but has no effect on predator density fluctuation. The 95% posterior predicted curves exceeds almost all of the bulk data values which demonstrate that the investigated models fit the data properly. Furthermore, stochastic models outperform deterministic models in terms of prediction efficiency. This discovery leads to the conclusion that the OU process can successfully handle noise in data that can arise from a variety of sources, such as model misspecification or measurement error, among others. The comparison of the presented models using the Information criteria revealed that differences are immaterial with the Holling type II model and Crowley Martin model to outperform using different criteria. The information criteria of the stochastic models, on the other hand, have over 100 units lower values than those of the deterministic models, which reflects the fact that stochasticity absorbs noise and allows for more accurate parameter estimation.

7.1.4 Vector born diseases

The mapping tools provided by the 'ggmap' library and presented here are suitable for visualizing critical risk measures on the map and monitoring the evolution of the VBDs. These mapping tools may propel the semi-automatic EWS developed in (Pergantas et al. 2021) to potentially aid in the monitoring and control of VBDs in a variety of settings. This kind of EWS model can enhance our ability to predict the risk of disease outbreaks while climatic conditions change. The developed methods can be adapted for usage in countries with similar vector-borne disease potential.

7.2 Future research

7.2.1 Demographic statistics in ecology

In terms of our approach to studying the reproductive potential of economically important species, it could be viewed as a broader tool when combined with models related to international trade and climatic change, because these models alert specialists to early detection strategies against invasive species and, as a result, their successful control. In addition, stochastic modelling of the variables (characteristics) of interest for *T. granarium* provides an assessment of the variability for such variables, thus providing plausible ranges for use in alternative conditions (e.g., temperature, relative humidity, commodity) for comparison with different but related species. Furthermore, the stochastic models used in this study allowed for model checking and characterization of the most suitable distribution for each component of the system, allowing for robust results and casting the two kinds of duration involved in this specific species within a larger taxon.

7.2.2 Modelling developmental rates

Regarding the study of developmental rates, among things for future research is to select consistently the most robust candidate between models given sufficiently many data samples, in a sensitivity analysis perspective. Moreover, the ADVI methods, can be extended so as to capture more sophisticated mean structures, like the ones we present in the current work. In addition, probability density that generates zeros in the Zero Inflated Inverse Gamma distribution can be modeled in more complex ways, such as using hyperparameters and hierarchical effects across temperature levels. Finally, R-packages that include the suggested models and perform the analysis presented in this Chapter are to be created.

7.2.3 Modelling predator-prey systems

In terms of predator-prey systems, future research may concentrate on developing fully predictive models with appropriate scoring rules. Another area of interest could be the use of optimal control techniques to solve predator-prey system constraints such as deterioration of potentially infected prey populations from disease or minimizing costs induced by biological control implementation.

7.2.4 Modelling VBDs

Concerning the augmented EWS model described to monitor VBDs can be used for optimizing the cost-effectiveness of distinct control measures and the integration of open geospatial and climatological data.

References

Aghdam, H. R., Fathipour, Y. and Kontodimas, D. (2011), 'Evaluation of non-linear models to describe development and fertility of codling moth at constant temperatures', *Entomologia Hellenica* **20**(1), 3–16.

Aitken, A. (1975), 'Insect travelers. i: Coleoptera. techn', *Bull* **31**, 191.

Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* **19**(6), 716–723.

Analytis, S. (1981), 'Relationship between temperature and development times in phytopathogenic fungus and in plant pests: a mathematical model', *Agric. Res. Athens* **5**, 133–159.

Arditi, R. and Ginzburg, L. R. (2012), *How species interact: altering the standard view on trophic ecology*, Oxford University Press.

Arthur, F., Ghimire, M., Myers, S. and Phillips, T. (2018), 'Evaluation of pyrethroid insecticides and insect growth regulators applied to different surfaces for control of trogoderma granarium (coleoptera: Dermestidae) the khapra beetle', *Journal of economic entomology* **111**(2), 612–619.

Athanassiou, C. G. and Kavallieratos, N. G. (2014), 'Evaluation of spinetoram and spinosad for control of prostephanus truncatus, rhyzopertha dominica, sitophilus oryzae, and tribolium confusum on stored grains under laboratory tests', *Journal of Pest Science* **87**(3), 469–483.

Athanassiou, C. G., Kavallieratos, N. G., Boukouvala, M. C., Mavroforos, M. E. and Kontodimas, D. C. (2015), 'Efficacy of alpha-cypermethrin and thiamethoxam against trogoderma granarium everts (coleoptera: Dermestidae) and tenebrio molitor l.(coleoptera: Tenebrionidae) on concrete', *Journal of Stored Products Research* **62**, 101–107.

Banks, H. (1977), 'Distribution and establishment of trogoderma granarium everts (coleoptera: Dermestidae): climatic and other influences', *Journal of Stored Products Research* **13**(4), 183–202.

Barber, A., Campbell, C., Crane, H., Lilley, R. and Tregidga, E. (2003), 'Biocontrol of two-spotted spider mite tetranychus urticae on dwarf hops by the phytoseiid mites phytoseiulus persimilis and neoseiulus californicus', *Biocontrol Science and Technology* **13**(3), 275–284.

Basedow, T., Hua, L. and Aggarwal, N. (2006), 'The infestation of vicia faba l.(fabaceae) by aphis fabae (scop.)(homoptera: Aphididae) under the influence of lamiaceae (ocimum basilicum l. and satureja hortensis l.)', *Journal of pest science* **79**(3), 149–154.

Beddington, J. R. (1975), 'Mutual interference between parasites or predators and its effect on searching efficiency', *The Journal of Animal Ecology* pp. 331–340.

Berryman, A. A. (1992), 'The orgins and evolution of predator-prey theory', *Ecology* **73**(5), 1530–1535.

Berven, K. A. (1990), 'Factors affecting population fluctuations in larval and adult stages of the wood frog (rana sylvatica)', *Ecology* **71**(4), 1599–1608.

Betancourt, M. (2017), 'A conceptual introduction to hamiltonian monte carlo', *arXiv preprint arXiv:1701.02434* .

Bieri, M., Baumgartner, J., Bianchi, G., Delucchi, V., Arx, R. v. et al. (1983), 'Development and fecundity of pea aphid (acyrthosiphon pisum harris) as affected by constant temperatures and by pea varieties.', *Mitteilungen der Schweizerischen Entomologischen Gesellschaft* **56**(1/2), 163–171.

Blei, D. M., Kucukelbir, A. and McAuliffe, J. D. (2017), 'Variational inference: A review for statisticians', *Journal of the American Statistical Association* **112**(518), 859–877.

Boukouvala, M. C., Kavallieratos, N. G., Athanassiou, C. G., Losic, D., Hadjiarapoglou, L. P. and Elemes, Y. (2017), 'Laboratory evaluation of five novel pyrrole derivatives as grain protectants against tribolium confusum and ephestia kuehniella larvae', *Journal of Pest Science* **90**(2), 569–585.

Bradshaw, C. J., Leroy, B., Bellard, C., Roiz, D., Albert, C., Fournier, A., Barbet-Massin, M., Salles, J.-M., Simard, F. and Courchamp, F. (2016), 'Massive yet grossly underestimated global costs of invasive insects', *Nature communications* **7**(1), 1–8.

Briere, J.-F. and Pracros, P. (1998), 'Comparison of temperature-dependent growth models with the development of lobesia botrana (lepidoptera: Tortricidae)', *Environmental Entomology* **27**(1), 94–101.

Briere, J.-F., Pracros, P., Le Roux, A.-Y. and Pierre, J.-s. (1999), 'A novel rate model of temperature-dependent development for arthropods', *Environmental Entomology* **28**(1), 22–29.

Briere, J., Pracros, P. and Stoeckel, J. (1998), 'Modeling development rate for predicting lobesia botrana (den. and schiff.) population dynamics', *IOBC WPRS Bulletin* **21**, 51–52.

Broufas, G. D. and Koveos, D. S. (2001), 'Development, survival and reproduction of euseius finlandicus (acari: Phytoseiidae) at different constant temperatures', *Experimental and applied acarology* **25**(6), 441–460.

Broufas, G., Pappas, M. and Koveos, D. (2007), 'Development, survival, and reproduction of the predatory mite kampimodromus aberrans (acari: Phytoseiidae) at different constant temperatures', *Environmental entomology* **36**(4), 657–665.

Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M. and West, G. B. (2004), 'Toward a metabolic theory of ecology', *Ecology* **85**(7), 1771–1789.

Buckland, S. T., Burnham, K. P. and Augustin, N. H. (1997), 'Model selection: an integral part of inference', *Biometrics* pp. 603–618.

Burg, S. (2014), *Hypobaric storage in food industry: advances in application and theory*, Elsevier.

Burges, H. (2008), 'Development of the khapra beetle, trogoderma granarium, in the lower part of its temperature range', *Journal of stored products research* **44**(1), 32–35.

Calzolari, M. (2016), 'Mosquito-borne diseases in europe: an emerging public health threat', *Reports in Parasitology* **5**, 1–12.
URL: *https://doi.org/10.2147/RIP.S56780*

Campbell, A., Frazer, B., Gilbert, N., Gutierrez, A. and Mackauer, M. (1974), 'Temperature requirements of some aphids and their parasites', *Journal of Applied Ecology* **11**, 431–438.

Carey, J. R. (1993), *Applied demography for biologists: with special emphasis on insects*, Oxford University Press.

Carey, J. R. (2001), 'Insect biodemography', *Annual review of entomology* **46**(1), 79–110.

Carey, J. R. and Vargas, R. I. (1985), 'Demographic analysis of insect mass rearing: a case study of three tephritids', *Journal of Economic Entomology* **78**(3), 523–527.

Carlin, B. P. and Chib, S. (1995), 'Bayesian model choice via markov chain monte carlo methods', *Journal of the Royal Statistical Society: Series B (Methodological)* **57**(3), 473–484.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P. and Riddell, A. (2017), 'Stan: A probabilistic programming language', *Journal of statistical software* **76**(1), 1–32.

Cohen, S. D., Hindmarsh, A. C. and Dubois, P. F. (1996), 'Cvode, a stiff/nonstiff ode solver in c', *Computers in physics* **10**(2), 138–143.

Colunga-Garcia, M., Haack, R., Magarey, R. and Borchert, D. (2013), 'Understanding trade pathways to target biosecurity surveillance', *NeoBiota* **18**, 103.

Compagnoni, A., Bibian, A. J., Ochocki, B. M., Rogers, H. S., Schultz, E. L., Sneck, M. E., Elderd, B. D., Iler, A. M., Inouye, D. W., Jacquemyn, H. et al. (2016), 'The effect of demographic correlations on the stochastic population dynamics of perennial plants', *Ecological Monographs* **86**(4), 480–494.

Cox, D. R. and Oakes, D. (1984), *Analysis of survival data*, Vol. 21, CRC press.

Crowley, P. H. and Martin, E. K. (1989), 'Functional responses and interference within and between year classes of a dragonfly population', *Journal of the North American Benthological Society* **8**(3), 211–221.

Damos, P. and Savopoulou-Soultani, M. (2012), 'Temperature-driven models for insect development and vital thermal requirements', *Psyche: A Journal of Entomology* **2012**, 13.

Day, W., Prokrym, D., Ellis, D., Chianese, R. et al. (1994), 'The known distribution of the predator propylea quatuordecimpunctata (coleoptera: Coccinellidae) in the united states, and thoughts on the origin of this species and five other exotic lady beetles in eastern north america.', *Entomological News* **105**(4), 244–256.

DeAngelis, D. L., Goldstein, R. and O'Neill, R. V. (1975), 'A model for tropic interaction', *Ecology* **56**(4), 881–892.

Degallier, N., Favier, C., Menkes, C., Lengaigne, M., Ramalho, W., Souza, R. and Boulanger, J.-P. (2010), 'Toward an early warning system for dengue prevention: modeling climate impact on dengue transmission', *Climatic Change* **98**, 581–592.
URL: *https://doi.org/10.1007/s10584-009-9747-3*

Dellaportas, P., Forster, J. J. and Ntzoufras, I. (2000), 'Bayesian variable selection using the gibbs sampler', *Biostatistics-Basel* **5**, 273–286.

Diane, L. (1992), 'Zero-inflated poisson regression, with an application to defects in manufacturing', *Technometrics* **34**(1), 1–14.
URL: *https://www.tandfonline.com/doi/abs/10.1080/00401706.1992.10485228*

Dixon, A. F., Honěk, A., Keil, P., Kotela, M. A. A., Šizling, A. L. and Jarošík, V. (2009), 'Relationship between the minimum and maximum temperature thresholds for development in insects', *Functional Ecology* **23**(2), 257–264.

Dormand, J. R. and Prince, P. J. (1980), 'A family of embedded runge-kutta formulae', *Journal of computational and applied mathematics* **6**(1), 19–26.

Dostalkova, I., Kindlmann, P. and Dixon, A. F. (2002), 'Are classical predator–prey models relevant to the real world?', *Journal of Theoretical Biology* **218**(3), 323–330.

Douma, J., Pautasso, M., Venette, R. C., Robinet, C., Hemerik, L., Mourits, M. C., Schans, J. and van der Werf, W. (2016), 'Pathway models for analysing and managing the introduction of alien plant pests an overview and categorization', *Ecological modelling* **339**, 58–67.

Efron, B. (1992), Bootstrap methods: another look at the jackknife, *in* 'Breakthroughs in statistics', Springer, pp. 569–593.

Engeman, R. M., Laborde, J. E., Constantin, B. U., Shwiff, S. A., Hall, P., Duffiney, A. and Luciano, F. (2010), 'The economic impacts to commercial farms from invasive monkeys in puerto rico', *Crop Protection* **29**(4), 401–405.

Engen, S., Lande, R., Sæther, B.-E. and Dobson, F. S. (2009), 'Reproductive value and the stochastic demography of age-structured populations', *The American Naturalist* **174**(6), 795–804.

EPPO (1981), 'Data sheets on quarantine organisms. eppo list a2. trogoderma granarium everts (coleoptera: Dermestidae).', *EPPO Bulletin* **11**, 1–9.

EPPO (2013), 'Pm 7/13 (2) trogoderma granarium', *EPPO Bulletin* **43**(3), 431–448.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/epp.12080*

EPPO (2018), 'Trogoderma granarium', *EPPO global data base* .
URL: *https://gd.eppo.int/taxon/TROGGA*

Eurostat (2020), 'Asylum applications (non-eu) in the eu-28 member states, 2008-2018', *Eurostat* **16**.
URL: *https://ec.europa.eu/eurostat/statistics-explained/index.php/Asylum_statistics*

Fernández, C. and Steel, M. (1998), 'On the dangers of modelling through continuous distribution: A bayesian perspective. bayesian statistics 6, jm bernardo, jo berger, ap dawid, and afm smith'.

Fotakis, E., Giantsis, I., Sierra, J., Tanti, F., Balaska, S., Mavridis, K. and Chaskopoulou, A. (2020), 'Population dynamics, pathogen detection and insecticide resistance of mosquito and sand fly in refugee camps, greece', *Infectious diseases of poverty* **9**, 30.

Friel, N. and Pettitt, A. N. (2008), 'Marginal likelihood estimation via power posteriors', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**(3), 589–607.

Frühwirth-Schnatter, S. (2004), 'Estimating marginal likelihoods for mixture and markov switching models using bridge sampling techniques', *The Econometrics Journal* **7**(1), 143–167.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2014), *Bayesian data analysis*, CRC press.

Gelman, A. and Meng, X.-L. (1998), 'Simulating normalizing constants: From importance sampling to bridge sampling to path sampling', *Statistical science* **13**(2), 163–185.

George, E. I. and McCulloch, R. E. (1997), 'Approaches for bayesian variable selection', *Statistica sinica* **7**(2), 339–373.

Ghimire, M. N., Myers, S. W., Arthur, F. H. and Phillips, T. W. (2017), 'Susceptibility of trogoderma granarium everts and trogoderma inclusum leconte (coleoptera: Dermestidae) to residual contact insecticides', *Journal of stored products research* **72**, 75–82.

Ghosh, S. K., Mukhopadhyay, P. and Lu, J.-C. J. (2006), 'Bayesian analysis of zero-inflated regression models', *Journal of Statistical planning and Inference* **136**(4), 1360–1375.

Green, P. J. (1995), 'Reversible jump markov chain monte carlo computation and bayesian model determination', *Biometrika* **82**(4), 711–732.

Gronau, Q. F., Singmann, H. and Wagenmakers, E.-J. (2020), 'bridgesampling: An R package for estimating normalizing constants', *Journal of Statistical Software* **92**(10), 1–29.

Hagstrum, D. W., Phillips, T. W. and Cuperus, G. (2012), 'Stored product protection', *Kansas State University, Manhattan, KS. KSRE Publ* pp. 297–304.

Halsey, L. G. (2019), 'The reign of the p-value is over: what alternative analyses could we employ to fill the power vacuum?', *Biology letters* **15**(5), 20190174.

Hare, M. P., Nunney, L., Schwartz, M. K., Ruzzante, D. E., Burford, M., Waples, R. S., Ruegg, K. and Palstra, F. (2011), 'Understanding and estimating effective population size for practical application in marine species management', *Conservation Biology* **25**(3), 438–449.

Hassell, M. and Varley, G. (1969), 'New inductive population model for insect parasites and its bearing on biological control', *Nature* **223**(5211), 1133–1137.

Hill, D. S. (2002), *Pests of stored foodstuffs and their control*, Springer Science and Business Media.

Hodek, I. and Honêk, A. (2013), *Ecology of coccinellidae*, Vol. 54, Springer Science and Business Media.

Holling, C. S. (1959*a*), 'The components of predation as revealed by a study of small-mammal predation of the european pine sawfly', *The Canadian Entomologist* **91**(5), 293–320.

Holling, C. S. (1959*b*), 'Some characteristics of simple types of predation and parasitism', *The Canadian Entomologist* **91**(7), 385–398.

Howe, R. (1965), 'A summary of estimates of optimal and minimal conditions for population increase of some stored products insects', *Journal of Stored Products Research* **1**(2), 177–184.

Huey, R. B. and Berrigan, D. (2001*a*), 'Temperature, demography, and ectotherm fitness', *The American Naturalist* **158**(2), 204–210.

Huey, R. and Berrigan, D. (2001*b*), 'Temperature, demography and ectotherm fitness', *American Naturalist* **158**, 204–210.

Huffaker, C., Messenger, P. and DeBach, P. (1971), The natural enemy component in natural control and the theory of biological control, *in* 'Biological Control', Huffaker, C.B., Plenum, New York, NY, p. 1 –67.

Hunter, C. M., Caswell, H., Runge, M. C., Regehr, E. V., Amstrup, S. C. and Stirling, I. (2010), 'Climate change threatens polar bear populations: a stochastic demographic analysis', *Ecology* **91**(10), 2883–2897.

HUREJ, M. and WERF, W. V. D. (1993), 'The influence of black bean aphid, aphis fabae scop., and its honeydew on the photosynthesis of sugar beet', *Annals of applied biology* **122**(2), 189–200.

Jalali, M., Tirry, L., Arbab, A. and De Clercq, P. (2010), 'Temperature-dependent development of the two-spotted ladybeetle, adalia bipunctata, on the green peach aphid, myzus persicae, and a factitious food under constant temperatures', *Journal of Insect Science* **10**, 124.

Jarošík, V., Honěk, A. and Dixon, A. F. (2002), 'Developmental rate isomorphy in insects and mites', *The American Naturalist* **160**(4), 497–510.

Jonzén, N., Pople, T., Knape, J. and Sköld, M. (2010), 'Stochastic demography and population dynamics in the red kangaroo macropus rufus', *Journal of Animal Ecology* **79**(1), 109–116.

Kahle, D. and Wickham, H. (2013), 'ggmap: Spatial visualization with ggplot2', *The R Journal* **5**, 144–161. URL.
URL: *http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf*

Kalbfleisch, J. D. and Prentice, R. L. (2011), *The statistical analysis of failure time data*, Vol. 360, John Wiley and Sons.

Kaplan, E. L. and Meier, P. (1958), 'Nonparametric estimation from incomplete observations', *Journal of the American statistical association* **53**(282), 457–481.

Kass, R. E. and Raftery, A. E. (1995), 'Bayes factors', *Journal of the american statistical association* **90**(430), 773–795.

Kavallieratos, N, G., Athanassiou, C, G., Diamantis, G, C., Gioukari, H, G. and Boukouvala, M, C. (2017a), 'Evaluation of six insecticides against adults and larvae of trogoderma granarium everts (coleoptera: Dermestidae) on wheat, barley, maize and rough rice', *Journal of Stored Products Research* **71**, 81–92.

Kavallieratos, N, G., Athanassiou, C, G., Guedes, R, N., Drempela, J, D. and Boukouvala, M. C. (2017b), 'Invader competition with local competitors: displacement or coexistence among the invasive khapra beetle, trogoderma granarium everts (coleoptera: Dermestidae), and two other major stored-grain beetles?', *Frontiers in plant science* **8**, 1837.

Kavallieratos, N. G., Athanassiou, C. G., Hatzikonstantinou, A. N. and Kavallieratou, H. N. (2011), 'Abiotic and biotic factors affect efficacy of chlorfenapyr for control of stored-product insect pests', *Journal of food protection* **74**(8), 1288–1299.

Kavallieratos, N. G. and Boukouvala, M. C. (2018), 'Efficacy of four insecticides on different types of storage bags for the management of trogoderma granarium everts (coleoptera: Dermestidae) adults and larvae', *Journal of stored products research* **78**, 50–58.

Kerkow, A., Wieland, R., Früh, L., Hölker, F., Jeschke, J., Werner, D. and Kampen, H. (2020), 'Can data from native mosquitoes support determining invasive species habitats? modelling the climatic niche of aedes japonicus japonicus (diptera, culicidae) in germany', *Parasitology Research* **119**, 31–42.
URL: *https://doi.org/10.1007/s00436-019-06513-5*

Kindlmann, P. and Dixon, A. F. (2001), 'When and why top-down regulation failsin arthropod predator-prey systems', *Basic and Applied Ecology* **2**(4), 333–340.

Kondakis, M., Demiris, N., Ntzoufras, I. and Papanikolaou, N. E. (2021a), 'Inference and model determination for temperature-driven non-linear ecological models', *arXiv preprint arXiv:2104.15043* .

Kondakis, M., Nikolaos, D. and Papanikolaou, N. E. (2021b), 'rdevel: R package for developmental rate inference', `https://https://github.com/mkondakis/rdevel`.

Kontodimas, D. C., Eliopoulos, P. A., Stathas, G. J. and Economou, L. P. (2004), 'Comparative temperature-dependent development of nephus includens (kirsch) and nephus bisignatus (boheman)(coleoptera: Coccinellidae) preying on planococcus citri (risso)(homoptera: Pseudococcidae): evaluation of a linear and various nonlinear models using specific criteria', *Environmental Entomology* **33**(1), 1–11.

Kontodimas, D. C., Milonas, P. G., Stathas, G. J., Papanikolaou, N. E., Skourti, A. and Matsinos, Y. G. (2008), 'Life table parameters of the aphid predators coccinella septempunctata, ceratomegilla undecimnotata and propylea quatuordecimpunctata (coleoptera: Coccinellidae)', *European Journal of Entomology* **105**(3), 427.

Kontodimas, D., Milonas, P., Stathas, G., Economou, L. and Kavallieratos, N. (2007), 'Life table parameters of the pseudococcid predators nephus includens and nephus bisignatus (coleoptera: Coccinelidae', *European Journal of Entomology* **104**, 407–415.

Kratina, P., Vos, M., Bateman, A. and Anholt, B. R. (2009), 'Functional responses modified by predator density', *Oecologia* **159**(2), 425–433.

Kriticos, D. J., Venette, R. C., Baker, R. H., Brunel, S., Koch, F. H., Rafoss, T., Van der Werf, W. and Worner, S. P. (2013), 'Invasive alien species in the food chain: Advanc-

ing risk assessment models to address climate change, economics and uncertainty', *NeoBiota. 18: 1-7.* **18**, 1–7.

Kuhn, K., Campbell-Lendrum, D., Haines, A. and Cox, J. (2005), *Using climate to predict infectious disease epidemics*, World Health Organization, Geneva.

Kuo, L. and Mallick, B. (1998), 'Variable selection for regression models', *Sankhyā: The Indian Journal of Statistics, Series B* **60**(1), 65–81.

Lactin, D. J., Holliday, N., Johnson, D. and Craigen, R. (1995), 'Improved rate model of temperature-dependent development by arthropods', *Environmental entomology* **24**(1), 68–75.

Lindgren, D. L. and Vincent, L. E. (1959), 'Biology and control of trogoderma granarium everts', *Journal of Economic Entomology* **52**(2), 312–319.

Lindgren, D., Vincent, L., Krohne, H. et al. (1955), 'The khapra beetle, trogoderma granarium everts', *Hilgardia* **24**(1), 1–36.

Lotka, A. J. (1925), *Elements of physical biology*, Williams and Wilkins.

Lowe, S., Browne, M., Boudjelas, S. and De Poorter, M. (2000), *100 of the world's worst invasive alien species: a selection from the global invasive species database*, Vol. 12, Invasive Species Specialist Group Auckland.

Lukaszyk, S. (2004), 'A new concept of probability metric and its applications in approximation of scattered data sets', *Computational Mechanics* **33**(4), 299–304.

Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000), 'Winbugs-a bayesian modelling framework: concepts, structure, and extensibility', *Statistics and computing* **10**(4), 325–337.

Macdonald, G. (1952), 'The analysis of equilibrium in malaria', *Tropical Diseases Bulletin* **49**, 813–829. PMID: 12995455.

Malesios, C., Demiris, N., Kalogeropoulos, K. and Ntzoufras, I. (2017), 'Bayesian epidemic models for spatially aggregated count data', *Statistics in medicine* **36**(20), 3216–3230.

Marras, S., Cucco, A., Antognarelli, F., Azzurro, E., Milazzo, M., Bariche, M., Butenschön, M., Kay, S., Di Bitetto, M., Quattrocchi, G. et al. (2015), 'Predicting future thermal habitat suitability of competing native and invasive fish species: from metabolic scope to oceanographic modelling', *Conservation physiology* **3**(1), cou059.

Medlock, J. M. and Leach, S. A. (2015), 'Effect of climate change on vector-borne disease risk in the uk', *The Lancet Infectious Diseases* **15**(6), 721–730.

Meng, X.-L. and Wong, W. H. (1996), 'Simulating ratios of normalizing constants via a simple identity: a theoretical exploration', *Statistica Sinica* **6**(4), 831–860.

Minakawa, N., Mutero, C., Githure, J., Beier, J. and Yan, G. (1999), 'Spatial distribution and habitat characterization of anopheline mosquito larvae in western kenya', *American Journal of Tropical Medicine and Hygiene* **61**, 1010–1016.

Müller, R., Reuss, F., Kendrovski, V. and Montag, D. (2019), Vector-borne diseases, *in* M. Marselle, J. Stadler, H. Korn, K. Irvine and A. Bonne, eds, 'Biodiversity and health in the face of climate change', Springer Nature, Switzerland, p. 67–90.

Neal, R. M. (2010), 'MCMC using Hamiltonian dynamics', *Handbook of Markov Chain Monte Carlo* **54**, 113–162.

Nedvěd, O. (2009), Chapter 252 - temperature, effects on development and growth, *in* V. H. Resh and R. T. Cardé, eds, 'Encyclopedia of Insects (Second Edition)', second edition edn, Academic Press, San Diego, pp. 990–993.
URL: *https://www.sciencedirect.com/science/article/pii/B9780123741448002617*

Neven, L. G. (2000), 'Physiological responses of insects to heat', *Postharvest Biology and Technology* **21**(1), 103–111.

Norris, A. L. and Kunz, T. H. (2012), 'Effects of solar radiation on animal thermoregulation', *Solar radiation* pp. 195–220.

Obrycki, J. J., Harwood, J. D., Kring, T. J. and O'Neil, R. J. (2009), 'Aphidophagy by coccinellidae: application of biological control in agroecosystems', *Biological control* **51**(2), 244–254.

Orr, H. A. (2009), 'Fitness and its role in evolutionary genetics', *Nature Reviews Genetics* **10**(8), 531–539.

Overstall, A. M. and Forster, J. J. (2010), 'Default bayesian model determination methods for generalised linear mixed models', *Computational Statistics and Data Analysis* **54**(12), 3269–3288.

Paini, D. R., Sheppard, A. W., Cook, D. C., De Barro, P. J., Worner, S. P. and Thomas, M. B. (2016), 'Global threat to agriculture from invasive species', *Proceedings of the National Academy of Sciences* **113**(27), 7575–7579.

Papanikolaou, N. E., Demiris, N., Milonas, P. G., Preston, S. and Kypraios, T. (2016b), 'Does mutual interference affect the feeding rate of aphidophagous coccinellids? a modeling perspective', *PloS one* **11**(1), e0146168.

Papanikolaou, N. E., Kavallieratos, N. G., Kondakis, M., Boukouvala, M. C., Nika, E. P. and Demiris, N. (2019), 'Elucidating fitness components of the invasive dermestid beetle trogoderma granarium combining deterministic and stochastic demography', *PloS one* **14**(2), e0212182.

Papanikolaou, N. E., Milonas, P. G., Kontodimas, D. C., Demiris, N. and Matsinos, Y. G. (2013), 'Temperature-dependent development, survival, longevity, and fecundity of propylea quatuordecimpunctata (coleoptera: Coccinellidae)', *Annals of the Entomological Society of America* **106**(2), 228–234.

Papanikolaou, N. E., Milonas, P. G., Kontodimas, D. C., Demiris, N. and Matsinos, Y. G. (2014), 'Life table analysis of propylea quatuordecimpunctata (coleoptera: Coccinellidae) at constant temperatures', *Annals of the Entomological Society of America* **107**(1), 158–162.

Papanikolaou, N. E., Williams, H., Demiris, N., Preston, S., Milonas, P. G. and Kypraios, T. (2016a), 'Bayesian inference and model choice for holling's disc equation: a case study on an insect predator-prey system', *Community Ecology* **17**(1), 71–78.

Peacock, E. R. et al. (1993), *Adults and larvae of hide, larder and carpet beetles and their relatives (Coleoptera: Dermestidae) and of derodontid beetles (Coleoptera: Derodontidae).*, Vol. 5, Natural History Museum.

Pejchar, L. and Mooney, H. A. (2009), 'Invasive species, ecosystem services and human well-being', *Trends in ecology and evolution* **24**(9), 497–504.

Pergantas, P., Papanikolaou, N. E., Malesios, C., Tsatsaris, A., Kondakis, M., Perganta, I., Tselentis, Y. and Demiris, N. (2021), 'Towards a semi-automatic early warning system for vector-borne diseases', *International Journal of Environmental Research and Public Health* **18**(4), 1823.

Pergantas, P., Tsatsaris, A., Malesios, C., Kriparakou, G., Demiris, N. and Tselentis, Y. (2017), 'A spatial predictive model for malaria resurgence in central greece integrating entomological, environmental and social data', *PLoS ONE* **12**, 0178836.
URL: *https://doi.org/10.1371/journal.pone.0178836*

Perrakis, K., Ntzoufras, I. and Tsionas, E. G. (2014), 'On the use of marginal posteriors in marginal likelihood estimation via importance sampling', *Computational Statistics and Data Analysis* **77**, 54–69.

Polanco, A., Brewster, C. and Miller, D. (2011), 'Population growth potential of the bed bug', *Cimex lectularius L.: A life* **2**, 173–185.

Pressat, R. (1985), 'Contribution des écarts de mortalité par âge à la différence des vies moyennes', *Population (french edition)* pp. 766–770.

Rahman, K. A., Sohi, G. S. and Sapra, A. (1945), 'Studies on stored grain pests in the punjab vi. biology of trogoderma granarium everts', *Indian J Agric Sci* **15**, 85–92.

Rosenzweig, M. L. (1969), 'Why the prey curve has a hump', *The American Naturalist* **103**(929), 81–87.

Rosenzweig, M. L. (1977), 'Aspects of biological exploitation', *The Quarterly Review of Biology* **52**(4), 371–380.

Rosenzweig, M. L. and MacArthur, R. H. (1963), 'Graphical representation and stability conditions of predator-prey interactions', *The American Naturalist* **97**(895), 209–223.

Sakamoto, Y., Ishiguro, M. and Kitagawa, G. (1986), 'Akaike information criterion statistics', *Dordrecht, The Netherlands: D. Reidel* **81**(10.5555), 26853.

Schaffner, F., Medlock, J. and Van Bortel, W. (2013), 'Public health significance of invasive mosquitoes in europe', *Clinical Microbiology and Infection* **19**, 685–92.

Schellhorn, N. A. and Andow, D. A. (2005), 'Response of coccinellids to their aphid prey at different spatial scales', *Population Ecology* **47**(1), 71–76.

Schwarz, G. et al. (1978), 'Estimating the dimension of a model', *The annals of statistics* **6**(2), 461–464.

Semenza, J. (2015), 'Prototype early warning systems for vector-borne diseases in europe', *International Journal of Environmental Research and Public Health* **12**, 6333–6351.

Skalski, G. T. and Gilliam, J. F. (2001), 'Functional responses with predator interference: viable alternatives to the holling type ii model', *Ecology* **82**(11), 3083–3092.

Smith, D., Dushoff, J. and McKenzie, F. (2004), 'The risk of a mosquito-borne infection in a heterogeneous environment', *PLOS Biology* **2**.

Smith, D. and McKenzie, F. (2004), 'Statics and dynamics of malaria infection in anopheles mosquitoes', *Malaria journal* **3**, 13.
URL: *https://doi.org/10.1186/1475-2875-3-13*

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A. (2002), 'Bayesian measures of model complexity and fit', *Journal of the royal statistical society: Series b (statistical methodology)* **64**(4), 583–639.

Stan Development Team (2021*a*), 'RStan: the R interface to Stan'. R package version 2.21.0.
URL: *http://mc-stan.org/*

Stan Development Team (2021*b*), 'The Stan Core Library'. Version 2.21.0.
URL: *http://mc-stan.org/*

Stark, J. D. and Banks, J. E. (2016), 'Developing demographic toxicity data: optimizing effort for predicting population outcomes', *PeerJ* **4**, e2067.

Sutherland, W. J. (1983), 'Aggregation and theideal free'distribution', *The Journal of Animal Ecology* pp. 821–828.

Team, R. D. C. (2021), 'R: A language and environment for statistical computing'.
URL: *http://www.R-project.org*

van der Have, T. M. (2008), Slaves to the Eyring equation?: Temperature dependence of life-history characters in developing ectotherms, PhD thesis, Wageningen University.

Venables, W. N. and Ripley, B. D. (2013), *Modern applied statistics with S-PLUS*, Springer Science and Business Media.

Volterra, V. (1926), 'Variazioni e fluttuazioni del numero d'individui in specie animali conviventi'.

Wagner, T., Olson, R. and Willers, J. (1991), 'Modeling arthropod development time', *Journal of Agricultural Entomology* **8**(4), 251–270.

Watanabe, S. (2010), 'Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory', *Journal of Machine Learning Research* **11**(Dec), 3571–3594.

Wei, L.-J. (1992), 'The accelerated failure time model: a useful alternative to the cox regression model in survival analysis', *Statistics in medicine* **11**(14-15), 1871–1879.

W.H.O. (2021), 'Vector-borne diseases'.
URL: *https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases*

Yamasita, W., Das, S. and Chapiro, G. (2018), 'Numerical modeling of mosquito population dynamics of aedes aegypti', *Parasites and Vectors* **11**, 245.
URL: *https://doi.org/10.1186/s13071-018-2829-1*

Yemshanov, D., Koch, F. H., Lu, B., Lyons, D. B., Prestemon, J. P., Scarr, T. and Koehler, K. (2014), 'There is no silver bullet: the value of diversification in planning invasive species surveillance', *Ecological Economics* **104**, 61–72.

Zeki, E., PAPANIKOlLAOU, N. E., Demiris, N. and Kontodimas, D. C. (2015), 'Comparison of the demographic parameters and survival of two phenotypes of harmonia axyridis (coleoptera: Coccinellidae).', *European Journal of Entomology* **112**(1).

Zhu, L., Sellers, K. F., Morris, D. S. and Shmueli, G. (2017), 'Bridging the gap: A generalized stochastic process for count data', *The American Statistician* **71**(1), 71–80.

Appendices

Appendix A

Power posterior for Gaussian and Inverse Gamma distribution

A.1 Power posterior for Gaussian and Inverse Gamma distribution

Following (Friel and Pettitt 2008) the power posterior in (3.14) includes the likelihood raised to the power of t $p(\theta_i|y)^t$.

In the the Gaussian case, the likelihood involved becomes:

$$p(y|\theta)^t = p^t\left(y|\mu,\sigma^2\right) = \frac{e^{-\frac{1}{2}\left(\sqrt{t}\cdot\frac{y-\mu}{\sigma}\right)^2}}{\sqrt{2\pi\sigma^2}} = p\left(y|\mu,\frac{\sigma^2}{t}\right)\cdot\frac{\sqrt{t}}{t}.$$

Inserting the current ecological model $r\left(T;\theta\right)$, the log of the power posterior is given by:

$$log\left(p(y|\theta)^t\right) = -\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\left(\sqrt{t}\cdot\frac{y-r\left(T;\theta\right)}{\sigma}\right)^2.$$

In the Inverse Gamma case, the likelihood involved becomes:

$$p(y|\theta)^t = p(y|\alpha,\beta)^t = \frac{\beta^{\alpha\cdot t}}{\Gamma^t\left(\alpha\right)}y^{-\alpha\cdot t-t}exp\left(-\frac{\beta\cdot t}{y}\right) =$$
$$\frac{\left(\beta\cdot t\right)^{\alpha\cdot t}}{\Gamma\left(\alpha\cdot t\right)}\frac{\Gamma\left(\alpha\cdot t\right)}{\Gamma^t\left(\alpha\right)\cdot\left(t\right)^{\alpha\cdot t}}y^{-\alpha\cdot t-1}y^{1-t}exp\left(-\frac{\beta\cdot t}{y}\right) =$$
$$InG\left(\alpha\cdot t,\beta\cdot t\right)\cdot\frac{\Gamma\left(\alpha\cdot t\right)}{\Gamma^t\left(\alpha\right)\cdot\left(t\right)^{\alpha\cdot t}}\cdot y^{1-t}.$$

Inserting the current ecological model $r\left(T;\theta\right)$, the log of the power posterior is given by:

$$log\left(p(y|\theta)^t\right) = \alpha\cdot t\cdot\left(log\left(\alpha-1\right)+log\left(r\left(T;\theta\right)\right)\right) -$$
$$t\cdot log\left(\Gamma\left(\alpha\right)\right) - t\cdot\left(\alpha+1\right)\cdot log\left(y\right) - \frac{\left(\alpha-1\right)\cdot\left(r\left(T;\theta\right)\right)\cdot t}{y}.$$

Appendix B

Variance approximations of the marginal likelihood approximations via Power posterior and via Importance sampling method

B.1 Estimation of Variance of Power posterior method

$$\hat{\sigma_y}^2 = Var\left\{logP\left(y|\theta\right)\right\} = Var\left\{\int_0^1 E\left\{logP(y|\theta)\right\}dt\right\} \simeq$$

$$\sum_{i=1}^{n-1}\frac{1}{4}Var\left\{E_{\theta|y,t_i}\left\{logP(y|\theta)\right\} + E_{\theta|y,t_i+1}\left\{logP(y|\theta)\right\}\right\}\left(t_{i+1}-t_i\right)^2 \simeq$$

$$\sum_{i=1}^{n-1}\left[\frac{sd^2{}_{\theta|y,t_i}}{2}\right]\cdot\left(t_{i+1}-t_i\right)^2$$

where $sd_{\theta|y,t_i}$ is the std error estimated at the t_i temperature

B.2 Estimation of Variance of Importance sampling method

The marginal likelihood estimate for the ith model is given by $m = p\left(y|M_i\right) = E_g\left\{\frac{q(\theta_i|y)}{g(\theta_i)}\right\} = \int\frac{q(\theta_i|y)}{g(\theta_i)}g\left(\theta_i\right)d\theta_i$ provided $g\left(\theta - i\right) > 0$ whenever $q\left(\theta_i|y\right) \neq 0$ where the density $q\left(\theta_i|\ y\right)$ is equal to $p\left(y|\theta_i,\phi_i\right)\cdot p\left(\theta_i,\phi_i\right)$ and the auxiliary importance function g used is the following: $g\left(\theta_i\right) = g\left(\theta_i,\phi_i\right) = p\left(\theta_i|\ y\right)p\left(\phi_i|\ y\right)$
Removing the ith model index we can evaluate the marginal likelihood via MC integration which gives the formula below:

$$\hat{z}_y = \hat{p}\left(y\right) = \frac{1}{M}\sum_{J=1}^{M}\frac{q\left(\theta^j|y\right)}{g\left(\theta^j\right)}$$

where $\theta^j, j = 1,2,3,\ldots,M$ are obtained from density: $g\left(\theta_i\right)$.

$$E_g\left(\hat{z}_y\right) = \frac{1}{M}\sum_{j=1}^{M}E_g\left\{\frac{q\left(\theta^j|y\right)}{g\left(\theta^j\right)}\right\} = p\left(y\right) = m.$$

Assuming that $Cov\left\{\frac{q\left(\theta^j|y\right)}{g(\theta^j)},\frac{q\left(\theta^i|y\right)}{g(\theta^i)}\right\} = 0,\ for\ i \neq\ j$

$V_g(\hat{z}_y) = V_g\left(\frac{1}{M}\sum_{j=1}^{M}\left\{\frac{q\left(\theta^j|y\right)}{g(\theta^j)}\right\}\right) =$

$\frac{1}{M^2}\sum_{j=1}^{M} V_g\left\{\frac{q\left(\theta^j|y\right)}{g(\theta^j)}\right\} + \frac{1}{M^2}\sum_{k,j=1}^{M} Cov_g\left\{\frac{q\left(\theta^k|y\right)}{g\left(\theta^k\right)}, \frac{q\left(\theta^j|y\right)}{g(\theta^j)}\right\} =$

$\frac{1}{M^2}\sum_{j=1}^{M} V_g\left\{\frac{q\left(\theta^j|y\right)}{g(\theta^j)}\right\} = \frac{1}{M}V_g\left\{\frac{q(\theta|y)}{g(\theta)}\right\} = \frac{1}{M}\left\{E_g\left\{\frac{q(\theta|y)}{g(\theta)}\right\}^2 - {E_g}^2\left\{\frac{q(\theta|y)}{g(\theta)}\right\}\right\} =$

$\frac{1}{M}\int_G\left\{\frac{q(\theta|y)}{g(\theta)}\right\}^2 g(\theta)\,d\theta - \frac{1}{M}m^2 \Rightarrow$

Also, using properties of expectation and the fact that $g(\theta)$ is proper

$V_g(\hat{z}_y) = \frac{1}{M}\int_G\left\{\frac{q(\theta|y)-m\cdot g(\theta)}{g(\theta)}\right\}^2 g(\theta)\,d\theta = \frac{\sigma_y^2}{M}$ The latest can be used to estimate the std error of the estimator $\hat{z}_y$ via the formula:

$$\hat{\sigma}_y = \sqrt{\frac{1}{K}\sum_{j=1}^{K}\left\{\frac{q\left(\theta^j|y\right)}{g\left(\theta^j\right)} - \hat{z}_y\right\}^2}$$

where $\theta^j, j = 1, 2, 3, \ldots, K$ are obtained from density $g(\theta_i)$.

Appendix C

R code to produce maps with risk measures distribution

C.1 R code to produce maps

```
# Load library ggmap
library(ggmap)
# Define geographical coordinates of Greece
greece<- c(left = 20.1500159034, bottom = 34.9199876979, right = 26.6041955909,
top = 41.8269046087)
# Define map "bbox" for different map types
height<- max(locs$lat) - min(locs$lat)
width<- max(locs$lon) - min(locs$lon)
sac_borders<- c(bottom = min(locs$lat) - 0.1 * height, top = max(locs$lat) + 0.1 *
height, left = min(locs$lon) - 0.1 * width, right = max(locs$lon) + 0.1 * width)
map<- get_stamenmap(sac_borders, zoom = 9, maptype = "toner-background")
map_terrain<- get_stamenmap(sac_borders, zoom = 9, maptype = "terrain")
# (Alternative map types: "toner-lite", “terrain-background”, “terrain-labels”, “terrain-
lines”, “toner”, “toner-2010”, “toner-2011”, “toner-background”, “toner-hybrid”, “toner-
labels”, “toner-lines”, “toner-lite”, “watercolor”)

    # Read data and create data frame with parameters required data<- read.table("data.txt",
header = TRUE)
locs<- subset(data, select = c("lat","lon","Ro","tau", "E"))

    # Create R_0 map with qmplot focused on the coordinates of the data points
qmplot(x = lon, y = lat, col= Ro, size= Ro, data = locs, maptype = "terrain", geom =
"point") +
scale_color_gradient(low = "blue", high = "red") +
ggtitle(label="Risk ($R_0$) in malaria resurgence in central Greece", subtitle = "year=2018")
+
xlab("lon") +
ylab("lat")+
```

```
theme(plot.title = element_text( size = 18, face = "bold",hjust = 0.5 ), plot.subtitle =
element_text(hjust = 0.5, face = "italic"))
# Create τ map with qmplot focused on the coordinates of the data points
qmplot(x = lon, y = lat, col= tau, size= tau, data = locs, maptype = "terrain", geom =
"point") +
scale_color_gradient(low = "blue", high = "red") +
ggtitle(label="Risk (tau) in malaria resurgence in central Greece", subtitle = "year=2018")
+
xlab("lon") +
ylab("lat")+
theme(plot.title = element_text( size = 18, face = "bold",hjust = 0.5 ), plot.subtitle =
element_text(hjust = 0.5, face = "italic"))
qmplot(x = lon, y = lat, col= E, size= E, data = locs, maptype = "terrain", geom =
"point") +
scale_color_gradient(low = "blue", high = "red") +
ggtitle(label="Risk (E) in malaria resurgence in central Greece", subtitle = "year=2018")
+
xlab("lon") +
ylab("lat")+
theme(plot.title = element_text( size = 18, face = "bold",hjust = 0.5 ), plot.subtitle =
element_text(hjust = 0.5, face = "italic"))

    # Create R0 map with predefined size
ggmap(map) + geom_point(data = locs, mapping = aes(x = lon, y = lat, col= Ro, size=
Ro)) +
scale_color_distiller(palette = "YlOrRd", direction = 1)+
ggtitle(label="Risk (R0) in malaria resurgence in central Greece", subtitle = "year=2018")
+
xlab("lon") +
ylab("lat")+
theme(plot.title = element_text( size = 18, face = "bold",hjust = 0.5 ), plot.subtitle =
element_text(hjust = 0.5, face = "italic"))
# Create τ map with predefined size
ggmap(map) +
geom_point(data = locs, mapping = aes(x = lon, y = lat, col= Ro, size= Ro)) +
scale_color_distiller(palette = "YlOrRd", direction = 1)+
ggtitle(label="Risk (τ) in malaria resurgence in central Greece", subtitle = "year=2018")
+
```

```
xlab("lon") +
ylab("lat")+
theme(plot.title = element_text( size = 18, face = "bold",hjust = 0.5 ), plot.subtitle =
element_text(hjust = 0.5, face = "italic"))
# Create E map with predefined size
ggmap(map) +
geom_point(data = locs, mapping = aes(x = lon, y = lat, col= E, size= E)) +
scale_color_distiller(palette = "YlOrRd", direction = 1)+
ggtitle(label="Risk (E) in malaria resurgence in central Greece", subtitle = "year=2018")
+
xlab("lon") +
ylab("lat")+
theme(plot.title = element_text( size = 18, face = "bold",hjust = 0.5 ), plot.subtitle =
element_text(hjust = 0.5, face = "italic"))

    # Create R0 heat map with points
ggmap(map_terrain)+
stat_density2d(data = locs, aes(x = lon, y = lat ,fill = stat(nlevel)), geom = "poly-
gon",bins=100) +
geom_point(data = locs, mapping =aes(x = lon, y = lat, col= Ro)) +
scale_color_gradient(low = "blue", high = "red")+
scale_fill_gradient(low = "yellow", high = "orange")+
ggtitle(label="Risk (R0) in malaria resurgence in central Greece", subtitle = "year=2018")
+
xlab("lon") +
ylab("lat")+
theme_grey(base_size = 10)+
theme(plot.title = element_text( size = 16, face = "bold",hjust = 0.5 ), plot.subtitle =
element_text(hjust = 0.5, face = "italic"))+
guides( size=FALSE, alpha=FALSE)+
labs(color="Ro",fill="density") # Create τ heat map with points
ggmap(map_terrain)+
stat_density2d(data = locs, aes(x = lon, y = lat ,fill = stat(nlevel)), geom = "poly-
gon",bins=100) +
geom_point(data = locs, mapping =aes(x = lon, y = lat, col= tau)) +
scale_color_gradient(low = "blue", high = "red")+
scale_fill_gradient(low = "yellow", high = "orange")+
ggtitle(label="Risk (τ) in malaria resurgence in central Greece", subtitle = "year=2018")
```

```
+
xlab("lon") +
ylab("lat")+
theme_grey(base_size = 10)+
theme(plot.title = element_text( size = 16, face = "bold",hjust = 0.5 ), plot.subtitle =
element_text(hjust = 0.5, face = "italic"))+
guides( size=FALSE, alpha=FALSE)+
labs(color="tau",fill="density")
# Create E heat map with points
ggmap(map_terrain)+
stat_density2d(data = locs, aes(x = lon, y = lat ,fill = stat(nlevel)), geom = "poly-
gon",bins=100) +
geom_point(data = locs, mapping =aes(x = lon, y = lat, col=E )) +
scale_color_gradient(low = "blue", high = "red")+
scale_fill_gradient(low = "yellow", high = "orange")+
ggtitle(label="Risk (E) in malaria resurgence in central Greece", subtitle = "year=2018")
+
xlab("lon") +
ylab("lat")+
theme_grey(base_size = 10)+
theme(plot.title = element_text( size = 16, face = "bold",hjust = 0.5 ), plot.subtitle =
element_text(hjust = 0.5, face = "italic"))+
guides( size=FALSE, alpha=FALSE)+
labs(color="E",fill="density")
```